NOVEL SINGLE NUCLEOTIDE POLYMORPHISMS FOR OLFACTORY RECEPTOR-LIKE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

RELATED APPLICATIONS

This application claims the benefit of priority to Provisional Application U.S. Serial No. 60/_____, filed September 20, 2001; USSN 09/777,789, filed February 6, 2001; and USSN 60/245,292, filed November 2, 2000; each of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD OF THE INVENTION

The invention generally relates to variants of proteins encoded by a cDNA. These variants are known by the term "single nucleotide polymorphisms" or "SNPs".

BACKGROUND OF THE INVENTION

Within the animal kingdom, odor detection is a universal tool used for social interaction, predation, and reproduction. Chemosensitivity in vertebrates is modulated by bipolar sensory neurons located in the olfactory epithelium, which extend a single, highly arborized dendrite into the mucosa while projecting axons to relay neurons within the olfactory bulb. The many ciliae on the neurons bear odorant (or olfactory) receptors (ORs), which cause depolarization and formation of action potentials upon contact with specific odorants. ORs may also function as axonal guidance molecules, a necessary function as the sensory neurons are normally renewed continuously through adulthood by underlying populations of basal cells.

The mammalian olfactory system is able to distinguish several thousand odorant molecules. Odorant receptors are believed to be encoded by an extremely large subfamily of G protein-coupled receptors. These receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals and possibly other chemosensing responses as well. The genes encoding these receptors are devoid of introns within their coding regions. Schurmans and co-workers cloned a member of this family of genes, OLFR1, from a genomic library by cross-hybridization with a gene fragment obtained by PCR. See Schurmans et al., 63(3) Cytogenet. Cell Genet. 200 (1993). By isotopic *in situ* hybridization, they mapped the gene to 17p13-p12 with a peak at band 17p13. A minor peak was detected on chromosome 3, with a maximum in the region 3q13-q21. After MspI digestion, a restriction fragment length

polymorphism (RFLP) was demonstrated. Using this in a study of 3 CEPH pedigrees, they demonstrated linkage with D17S126 at 17pter-p12; maximum lod = 3.6 at theta = 0.0. Used as a probe on Southern blots under moderately stringent conditions, the cDNA hybridized to at least 3 closely related genes. Ben-Arie and colleagues cloned 16 human OLFR genes, all from 17p13.3. See Ben-Arie et al., 3(2) Hum. Mol. Genet. 229(1994). The intronless coding regions are mapped to a 350-kb contiguous cluster, with an average intergenic separation of 15 kb. The OLFR genes in the cluster belong to 4 different gene subfamilies, displaying as much sequence variability as any randomly selected group of OLFRs. This suggested that the cluster may be one of several copies of an ancestral OLFR gene repertoire whose existence may have predated the divergence of mammals. Localization to 17p13.3 was performed by fluorescence *in situ* hybridization as well as by somatic cell hybrid mapping.

Previously, OR genes cloned in different species were from disparate locations in the respective genomes. The human OR genes, on the other hand, lack introns and may be segregated into four different gene subfamilies, displaying great sequence variability. These genes are primarily expressed in olfactory epithelium, but may be found in other chemoresponsive cells and tissues as well.

Blache and co-workers used polymerase chain reaction (PCR) to clone an intronless cDNA encoding a new member (named OL2) of the G protein-coupled receptor superfamily. See Blache et al., 242(3) Biochem. Biophys. Res. Commun. 669 (1998). The coding region of the rat OL2 receptor gene predicts a seven transmembrane domain receptor of 315 amino acids. OL2 has 46.4 percent amino acid identity with OL1, an olfactory receptor expressed in the developing rat heart, and slightly lower percent identities with several other olfactory receptors. PCR analysis reveals that the transcript is present mainly in the rat spleen and in a mouse insulin-secreting cell line (MIN6). No correlation was found between the tissue distribution of OL2 and that of the olfaction-related GTP-binding protein Golf alpha subunit. These findings suggest a role for this
new hypothetical G-protein coupled receptor and for its still unknown ligand in the spleen and in the insulin-secreting beta cells.

Olfactory loss may be induced by trauma or by neoplastic growths in the olfactory neuroepithelium. There is currently no treatment available that effectively restores olfaction in the case of sensorineural olfactory losses. See Harrison's Principles of Internal Medicine, 14th Ed., Fauci, A.S. et al., Eds., McGraw-Hill, New York, p. 173 (1998). There thus remains a need for effective treatment to restore olfaction in pathologies related to neural olfactory loss.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel polynucleotide sequences encoding novel polypeptides.

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule that includes the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 and 13, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide that includes the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12 and 14. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule. Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a SEQ ID NO: 1, 3, 5, 7, 9, 11 and 13 nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified SEQ ID NO: 2, 4, 6, 8, 10, 12 and 14 polypeptide, *e.g.*, any of the polypeptides encoded by a SEQ ID NO: 1, 3, 5, 7, 9, 11 and 13 nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide and a pharmaceutically acceptable carrier or diluent.

In still a further aspect, the invention provides an antibody that binds specifically to a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including NOVX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide by providing a cell containing an encoding nucleic acid, *e.g.*, a vector that includes a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid, and culturing the cell under

conditions sufficient to express the SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide encoded by the nucleic acid. The expressed SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide or SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide by contacting a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide with a compound and determining whether the activity of that polypeptide is modified.

The invention is also directed to compounds that modulate SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide activity identified by contacting a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide with the compound and determining whether the compound modifies activity of that polypeptide, binds to that polypeptide, or binds to a nucleic acid molecule encoding that polypeptide.

In another aspect, the invention provides a method of determining the presence of or predisposition of a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in the subject sample. The amount of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in the subject sample is then compared to the amount of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in a control sample. An alteration in the amount of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in the subject protein sample relative to the amount of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide is detected using a complementary antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13-associated disorder in a subject. The

method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid in the subject nucleic acid sample. The amount of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid sample in the subject nucleic acid is then compared to the amount of a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid in a control sample. An alteration in the amount of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid in the sample relative to the amount of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 in the control sample indicates the subject has a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or delaying a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid-associated disorder, or a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 nucleic acid, a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide, or an antibody complementary to a SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14 polypeptide in an amount sufficient to treat, prevent, or delay the associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

Olfactory receptors (ORs) are the largest family of G-protein-coupled receptors (GPCRs) and belong to the first family (Class A) of GPCRs, along with catecholamine receptors and opsins. The OR family contains over 1,000 members that traverse the phylogenetic spectrum from *C. elegans* to mammals. ORs most likely emerged from prototypic GPCRs several times independently, extending the structural diversity necessary both within and between species in order to differentiate the multitude of ligands. Individual olfactory sensory neurons are predicted

to express a single, or at most a few, ORs. All ORs are believed to contain seven α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. The pocket of OR ligand binding is expected to be between the second and sixth transmembrane domains of the proteins. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%, and genes greater than 80% identical to one another at the amino acid level are considered to belong to the same subfamily.

Since the first ORs were cloned in 1991, outstanding progress has been made into their mechanisms of action and potential dysregulation during disease and disorder. It is understood that some human diseases result from rare mutations within GPCRs. Drug discovery avenues could be used to produce highly specific compounds on the basis of minute structural differences of OR subtypes, which are now being appreciated with *in vivo* manipulation of OR levels in transgenic and knock-out animals. Furthermore, due to the intracellular homogeneity and ligand specificity of ORs, renewal of specific odorant-sensing neurons lost in disease or disorder is possible by the introduction of individual ORs into basal cells. Additionally, new therapeutic strategies may be elucidated by further study of so-called orphan receptors, whose ligand(s) remain to be discovered.

OR proteins bind odorant ligands and transmit a G-protein-mediated intracellular signal, resulting in generation of an action potential. The accumulation of DNA sequences of hundreds of OR genes provides an opportunity to predict features related to their structure, function and evolutionary diversification. *See* Pilpel Y. et al., 33 Essays Biochem 93-104 (1993). The OR repertoire has evolved a variable ligand-binding site that ascertains recognition of multiple odorants, coupled to constant regions that mediate the cAMP-mediated signal transduction. The cellular second messenger underlies the responses to diverse odorants through the direct gating of olfactory-specific cation channels. This situation necessitates a mechanism of cellular exclusion, whereby each sensory neuron expresses only one receptor type, which in turn influences axonal projections. A 'synaptic image' of the OR repertoire thus encodes the detected odorant in the central nervous system.

The ability to distinguish different odors depends on a large number of different odorant receptors (ORs). ORs are expressed by nasal olfactory sensory neurons, and each neuron expresses only 1 allele of a single OR gene. In the nose, different sets of ORs are expressed in distinct spatial zones. Neurons that express the same OR gene are located in the same zone; however, in that zone they are randomly interspersed with neurons expressing other ORs. When

the cell chooses an OR gene for expression, it may be restricted to a specific zonal gene set, but it may select from that set by a stochastic mechanism. Proposed models of OR gene choice fall into 2 classes: locus-dependent and locus-independent. Locus-dependent models posit that OR genes are clustered in the genome, perhaps with members of different zonal gene sets clustered at distinct loci. In contrast, locus-independent models do not require that OR genes be clustered. OR genes have been mapped to 11 different regions on 7 chromosomes. These loci lie within paralogous chromosomal regions that appear to have arisen by duplications of large chromosomal domains followed by extensive gene duplication and divergence. Studies have shown that OR genes expressed in the same zone map to numerous loci; moreover, a single locus can contain genes expressed in different zones. These findings raised the possibility that OR gene choice is locus-independent or involved consecutive stochastic choices.

Issel-Tarver and Rine characterized 4 members of the canine olfactory receptor gene family (Issel-Tarver and Rine, "Organization and expression of canine olfactory genes," 93(20) PNAS, USA 10897-902 (October 1, 1996)). The 4 subfamilies comprised genes expressed exclusively in olfactory epithelium. Analysis of large DNA fragments using Southern blots of pulsed field gels indicated that subfamily members were clustered together, and that two of the subfamilies were closely linked in the dog genome. Analysis of the four olfactory receptor gene subfamilies in 26 breeds of dog provided evidence that the number of genes per subfamily was stable in spite of differential selection on the basis of olfactory acuity in scent hounds, sight hounds, and toy breeds.

Issel-Tarver and Rine performed a comparative study of four subfamilies of olfactory receptor genes first identified in the dog to assess changes in the gene family during mammalian evolution, and to begin linking the dog genetic map to that of humans (Issel-Tarver and Rine, "The evolution of mammalian olfactory receptor genes," 145(1) Genetics 185-95 (January, 1997)). These four families were designated by them OLF1, OLF2, OLF3, and OLF4 in the canine genome. The subfamilies represented by these four genes range in size from 2 to 20 genes. They are all expressed in canine olfactory epithelium but were not detectably expressed in canine lung, liver, ovary, spleen, testis, or tongue. The OLF1 and OLF2 subfamilies are tightly linked in the dog genome and also in the human genome. The smallest family is represented by the canine OLF1 gene. Using dog gene probes individually to hybridize to Southern blots of genomic DNA from 24 somatic cell hybrid lines. They showed that the human homologous OLF1 subfamily maps to human chromosome 11. The human gene with the strongest similarity to the canine OLF2 gene also mapped to chromosome 11. Both members of the human subfamily

that hybridized to canine OLF3 were located on chromosome 7. It was difficult to determine to which chromosome or chromosomes the human genes that hybridized to the canine OLF4 probe mapped. This subfamily is large in mouse and hamster as well as human, so the rodent background largely obscured the human cross-hybridizing bands. It was possible, however, to discern some human-specific bands in blots corresponding to human chromosome 19. They refined the mapping of the human OLF1 homolog by hybridization to YACs that map to 11q11. In dogs, the OLF1 and OLF2 subfamilies are within 45 kb of one another (Issel-Tarver and Rine (1996)).

Issel-Tarver and Rine demonstrated that in the human OLF1 and OLF2 homologs are likewise closely linked (Id.). By studying YACs, Issel-Tarver and Rine found that the human OLF3 homolog maps to 7q35. A chromosome 19-specific cosmid library was screened by hybridization with the canine OLF4 gene probe, and clones that hybridized strongly to the probe even at high stringency were localized to 19p13.1 and 19p13.2. These clones accounted, however, for a small fraction of the homologous human bands.

Rouquier et al. demonstrated that members of the olfactory receptor gene family are distributed on all but a few human chromosomes (Rouquier et al., 7(9) Human Molecular Genetics 1337-45 (Sept., 1998)). Through fluorescence *in situ* hybridization analysis, they showed that OR sequences reside at more than 25 locations in the human genome. Their distribution was biased for terminal bands of chromosome arms. Flow-sorted chromosomes were used to isolate 87 OR sequences derived from 16 chromosomes. Their sequence relationships indicated the inter- and intrachromosomal duplications responsible for OR family expansion. Rouquier et al. determined that the human genome has accumulated a striking number of dysfunctional copies: 72% of these sequences were found to be pseudogenes (Id.). ORF-containing sequences predominate on chromosomes 7, 16, and 17.

Trask et al. characterized a subtelomeric DNA duplication that provided insight into the variability, complexity, and evolutionary history of that unusual region of the human genome, the telomere (Trask et al., 7(13) Human Molec. Genetics 2007-20 (Dec. 12, 1998)). Using a DNA segment cloned from chromosome 19, they demonstrated that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Three chromosomes appeared to have contained the sequence before humans migrated around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in chimpanzee and gorilla, that site being nonorthologous to any of the locations in the human genome. Three new members of the olfactory receptor (OR) gene family were found to

be duplicated within this large segment of DNA, which was found to be present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. From its sequence, one of the OR genes in this duplicated block appeared to be potentially functional. The findings raised the possibility that functional diversity in the OR family is generated in part through duplications and interchromosomal rearrangements of the DNA near human telomeres.

Mombaerts reviewed the molecular biology of the odorant receptor (OR) genes in vertebrates (Mombaerts, 286(5440) Science 707-711 (Review) (1999)). Buck and Axel discovered this large family of genes encoding putative odorant receptor genes (Buck and Axel, 65(1) Cell 175-87 (1991)). Zhao et al. provided functional proof that one OR gene encodes a receptor for odorants (Zhao et al., 279(5348) Science 237-47 (1998)). The isolation of OR genes from the rat by Buck and Axel was based on three assumptions (Ibid.). First, ORs are likely G protein-coupled receptors, which characteristically are 7-transmembrane proteins. Second, ORs are likely members of a multigene family of considerable size, because an immense number of chemicals with vastly different structures can be detected and discriminated by the vertebrate olfactory system. Third, ORs are likely expressed selectively in olfactory sensory neurons. Ben-Arie et al. (1994) focused attention on a cluster of human OR genes on 17p, to which the first human OR gene, OR1D2, had been mapped by Schurmans et al. (Schurmans et al., 63(3) Cytogenet. Cell Genetics 200-204 (1993)). According to Mombaerts, the sequences of more than 150 human OR clones had been reported (Mombaerts, 286(5440) Science 707-711 (Review) (1999)). The human OR genes differ markedly from their counterparts in other species by their high frequency of pseudogenes, except the testicular OR genes. Research showed that individual olfactory sensory neurons express a small subset of the OR repertoire. In rat and mouse, axons of neurons expressing the same OR converge onto defined glomeruli in the olfactory bulb.

OR proteins bind odorant ligands and transmit a G-protein-mediated intracellular signal, resulting in generation of an action potential. The accumulation of DNA sequences of hundreds of OR genes provides an opportunity to predict features related to their structure, function and evolutionary diversification. The OR repertoire has evolved a variable ligand-binding site that ascertains recognition of multiple odorants, coupled to constant regions that mediate the cAMP-mediated signal transduction. The cellular second messenger underlies the responses to diverse odorants through the direct gating of olfactory-specific cation channels. This situation necessitates a mechanism of cellular exclusion, whereby each sensory neuron expresses only one receptor type, which in turn influences axonal projections. A 'synaptic image' of the OR

repertoire thus encodes the detected odorant in the central nervous system. (*See* Pilpel et al., 9(4) Curr. Opin. Neurobiol. 419-26 (1999)).

The odorant-induced Ca(2+) increase inside the cilia of vertebrate olfactory sensory neurons controls both excitation and adaptation. The increase in the internal concentration of Ca(2+) in the cilia has recently been visualized directly and has been attributed to Ca(2+) entry through cAMP-gated channels. These recent results have made it possible to further characterize Ca(2+)'s activities in olfactory neurons. Ca(2+) exerts its excitatory role by directly activating Cl(-) channels. Given the unusually high concentration of ciliary Cl(-), Ca(2+)'s activation of Cl(-) channels causes an efflux of Cl(-) from the cilia, contributing high-gain and low-noise amplification to the olfactory neuron depolarization. Moreover, in combination with calmodulin, Ca(2+) mediates odorant adaptation by desensitizing cAMP-gated channels. The restoration of the Ca(2+) concentration to basal levels occurs via a Na(+)/Ca(2+) exchanger, which extrudes Ca(2+) from the olfactory cilia. (*See* Menini, 45(3) Cell Mol Biol (Noisy-le-grand) 285-91 (1999)).

The olfactory epithelium is unique in the mammalian nervous system as it is a site of continual neurogenesis. Constant turnover of primary sensory neurons in the periphery results in continuous remodeling of neuronal circuits and synapses in the olfactory bulb throughout life. Most of the specific mechanisms and factors that control and modulate this process are not known. Recent studies suggest that growth factors, and their receptors, may play a crucial role in the development and continuous regeneration of olfactory neurons, *i.e.* particularly in neuronal proliferation, neurite outgrowth, fasciculation and synapse formation of the olfactory system. The potential role of the following factors and their receptors in different species are reviewed: Nerve growth factor (NGF); insulin-like growth factors (IGFs); fibroblast growth factors (FGFs); epidermal growth factor (EGF); transforming growth factor alpha (TGF alpha); amphiregulin (AR) and transforming growth factors beta (TGFs beta). (*See* Plendl et al., 65(7) Biochemistry 824-33 (2000).

An important recent advance in the understanding of odor adaptation has come from the discovery that complex mechanisms of odor adaptation already take place at the earliest stage of the olfactory system, in the olfactory cilia. At least two rapid forms and one persistent form of odor adaptation coexist in vertebrate olfactory receptor neurons. These three different adaptation phenomena can be dissected on the basis of their different onset and recovery time courses and their pharmacological properties, indicating that they are controlled, at least in part, by separate molecular mechanisms. Evidence is provided for the involvement of distinct molecular steps in

these forms of odor adaptation, including Ca(2+) entry through cyclic nucleotide-gated (CNG) channels, Ca(2+)-dependent CNG channel modulation, Ca(2+)/calmodulin kinase II-dependent attenuation of adenylyl cyclase, and the activity of the carbon monoxide/cyclic GMP second messenger system. Identification of these molecular steps may help to elucidate how the olfactory system extracts temporal and intensity information and to which extent odor perception is influenced by the different mechanisms underlying adaptation. (*See* Zufall et al., 126(1) Comp. Biochem. Physiol. and Mol. Integr. Physiol.17-32 (2000)) .

Since the discovery of odorant-activated adenylate cyclase in the olfactory receptor cilia, research into the olfactory perception of vertebrates has rapidly expanded. Recent studies have shown how the odor discrimination starts at the receptor level: each of 700-1000 types of the olfactory neurons in the neural olfactory epithelium contains a single type of odor receptor protein. Although the receptors have relatively low specific affinities for odorants, excitation of different types of receptors forms an excitation pattern specific to each odorant in the glomerular layer of the olfactory bulb. It was demonstrated that adenosine 3',5'-cyclic monophosphate (cAMP) is very likely the sole second messenger for olfactory transduction. It was also demonstrated that the affinity of the cyclic nucleotide-gated channel for cAMP regulated by Ca(2+)/calmodulin is solely responsible for the adaptation of the cell. However, many other regulatory components were found in the transduction cascade. Regulated by Ca(2+) and/or the protein-phosphorylation, many of them may serve for the adaptation of the cell, probably on a longer time scale. It may be important to consider the resensitization as a part of this adaptation, as well as to collect kinetic data of each reaction to gain further insight into the olfactory mechanism. (*See* Nakamura, 193(1) J. Soc. Biol. 35-40 (1999) (PMID: 10908849, UI: 20371128)).

The olfactory epithelium (OE) of the mammal is uniquely suited as a model system for studying how neurogenesis and cell death interact to regulate neuron number during development and regeneration. To identify factors regulating neurogenesis and neuronal death in the OE, and to determine the mechanisms by which these factors act, investigators studied OE using two major experimental paradigms: tissue culture of OE; and ablation of the olfactory bulb or severing the olfactory nerve in adult animals, procedures that induce cell death and a subsequent surge of neurogenesis in the OE *in vivo*. These studies characterized the cellular stages in the olfactory receptor neuron (ORN) lineage, leading to the realization that at least three distinct stages of proliferating neuronal precursor cells are employed in generating ORNs. The identification of a number of factors that act to regulate proliferation and survival of ORNs and

their precursors suggests that these multiple developmental stages may serve as control points at which cell number is regulated by extrinsic factors. *In vivo* surgical studies, which have shown that all cell types in the neuronal lineage of the OE undergo apoptotic cell death, support this idea. These studies, and the possible coregulation of neuronal birth and apoptosis in the OE, are discussed. (*See* Calof et al., 196 Ciba Found. Symp. 188-210 (1996) (PMID: 8727984, UI: 96284837)).

To identify factors regulating neurogenesis and neuronal death in mammals and to determine the mechanisms by which these factors act, researchers studied mouse olfactory epithelium using two different experimental paradigms: tissue culture of olfactory epithelium purified from mouse embryos; and ablation of the olfactory bulb in adult mice, a procedure that induces olfactory receptor neuron (ORN) death and neurogenesis *in vivo*. Studies of olfactory epithelium cultures have allowed the characterization of the cellular stages in olfactory neurogenesis and to identify factors regulating proliferation and differentiation of precursor cells in the ORN lineage. Studies of adult olfactory epithelium determined that all cell types in this lineage-proliferating neuronal precursors, immature ORNs and mature ORNs-undergo cell death following olfactory bulb ablation and that this death has characteristics of programmed cell death or apoptosis. *In vitro* studies have confirmed that neuronal cells of the olfactory epithelium undergo apoptotic death and have permitted identification of several polypeptide growth factors that promote survival of a fraction of ORNs. Using this information, researchers have begun to explore whether these factors, as well as genes known to play crucial roles in cell death in other systems, function to regulate apoptosis and neuronal regeneration in the adult olfactory epithelium following lesion-induced ORN death. (PMID: 8866135, UI: 97019661).

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel NOV1, NOV2, NOV3, NOV4, NOV5, NOV6 and NOV7 nucleic acid sequences and their polypeptides. The NOV1, NOV2, NOV3, NOV4, NOV5, NOV6 and NOV7 sequences are collectively referred to as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides. Example 1 provides a description of how the novel nucleic acids were identified.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	AL135841_B	1	2	OR GPCR
2	AL135841_B	3	4	OR GPCR
3	AL135841_A	5	6	OR GPCR
4	CG54212-01	7	8	OR GPCR
5	Variant 13019736	9	10	OR GPCR
6	CG53482-01	11	12	OR GPCR
7	Variant 13373788	13	14	OR GPCR

Where OR GPCR is an odorant receptor of the G-protein coupled-receptor family.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

For example, NOV1-7 are homologous to members of the odorant receptor (OR) family of the human G-protein coupled receptor (GPCR) superfamily of proteins, as shown in Table 1. Thus, the NOV1-7 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders of olfactory loss, *e.g.*, trauma, HIV illness, neoplastic growth and neurological disorders *e.g.* Parkinson's disease and Alzheimer's disease.

In addition, the present invention also discloses novel variants for the olfactory receptor-like protein encoded by a cDNA, and their utility as markers for genetic traits involved in cardiovascular, endocrine, metabolic, neurologic, psychiatric, autoimmune, inflammatory, and oncologic diseases.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell motility, cell proliferation and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A NOV1 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV1 nucleic acid and its encoded polypeptide includes the sequences shown in Table 2. The disclosed nucleic acid (SEQ ID NO: 1) is 1,050 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 59-61 and ends with a TAA stop codon at nucleotides 995-997. The representative ORF encodes a 312 amino acid polypeptide (SEQ ID NO: 2). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 1.

TABLE 2

CCCTGTACCCTCTCTCCTTCCATCCCAGCTGTGGACCATCTCTTCAGAACTCTGCAGC**ATG**GAGCCGCTCAACAGAA
CAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCA
GCCATGTACCTGGTGACCCTCCTGGGGAACACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCC
CGTGTACTTCTTCCTGGGCAACCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCC
ACCTCCTGTCATCCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCC
ACGGAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCACGTGCT
CATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGTCGGTGACTGAGATGG
TCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCCTGGCAGTGCTGAAG
CTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGCCTGTACCCCTGGC
ATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCT
TCTCCACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAG
AGTAAGGAAGCCCACATCTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCAT
CTACAGCCTGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGG**TGA**GGGA
GGGCGGGGCTCTGTACAGACGCAGGTCTCAGGTTAGTAGCTGAGGCCAT (SEQ ID NO. 1)

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLSTLDICYTPTF
VPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLL
KSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSA
ARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRS
RASR (SEQ ID NO. 2)

The NOV1 nucleic acid sequence has homology (85% identity) with the mouse olfactory receptor gene cluster OR17 and OR6 (OLF) (SEQ ID NO: 15) (GenBank Accession No:AJ251155), as shown in Table 3. Also, the NOV1 polypeptide has homology (82% identity) to the mouse olfactory receptor 71 (OLF) (SEQ ID NO: 16) (GenBank Accession No: NP_062359), as is shown in Table 4.

Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. (See Dryer and Berghard, 20 Trends in Pharmacological Sciences 413 (1999)).

OR proteins have seven transmembrane α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence aligment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, NOV1 is predicted to have a seven transmembrane region and is similar in that region to representative olfactory receptor GPCRs of monkey (SEQ ID NO:17) (GenBank Accession No:AAF40368), mouse (SEQ ID NO:18) (GenBank Accession No:CAB55597), rat (SEQ ID NO:19) (GenBank Accession No:S29711), and human (SEQ ID NO:20) (GenBank Accession No:CAB96728), as shown in Table 5.

TABLE 3

```
NOV1:  99  tgaaaggattttctggctacccagccctggagcatctgctcttccctctgtgctcagcca 158
           |||| |||||||||||||||| ||||| |||| || ||||| |||||||||||||| ||
OLF: 6102  tgaagggattttctggctacccggccctcgagcggctactctttcctctgtgctcagtca 6161

NOV1:159  tgtacctggtgaccctcctggggaacacagccatcatggcggtgagcgtgctagatatcc 218
          |||||||||||||| || |||||||||||||||||| |||||||||||| || | ||| ||
OLF: 6162 tgtacctggtgactctgctggggaacacagccatcgtggcggtgagcatgttggatgccc 6221

NOV1:219  acctgcacacgcccgtgtacttcttcctgggcaacctctctaccctggacatctgctaca 278
          |||||||||||||| ||||||| |||||||| ||||| || |   ||||||||||||||||
OLF: 6222 gcctgcacacgcccatgtacttttcctgggtaacctttccattttggacatctgctaca 6281

NOV1:279  cgcccacctttgtgcctctgatgctggtccacctcctgtcatcccggaagaccatctcct 338
          |  | || ||||| || |||||||||||||||||||||| ||||||||||||||||||
OLF: 6282 catctacttttgtaccccctgatgctggtccacctcctgtcgtcccggaagaccatctcct 6341

NOV1:339  ttgctgtctgtgccatccagatgtgtctgagcctgtccacgggctccacggagtgcctgc 398
          || | | ||||||| |||||||||||||||||||| |||||||||||||| ||||||||||
OLF: 6342 ttacgggctgtgccgtccagatgtgtctgagcctctccacgggctccaccgagtgcctgc 6401

NOV1:399  tactggccatcacggcctatgaccgctacctggccatctgccagccactcaggtaccacg 458
          |  ||||| ||| |||||||||||||||| ||||||| |||||||||||||||||| ||
OLF: 6402 tgttggccgtcatggcctatgaccgctacttggccatttgccagccactcaggtaccccg 6461

NOV1:459  tgctcatgagccaccggctctgcgtgctgctgatgggagctgcctgggtcctctgcctcc 518
          |||||||||||||| |||||||| || |||||  |||||  ||||||| |||||||||
OLF: 6462 tgctcatgagccacaggctctgcctgatgctggcaggagcctcctgggtgctctgcctct 6521

NOV1:519  tcaagtcggtgactgagatggtcatctccatgaggctgcccttctgtggccaccacgtgg 578
          ||||||| ||| | |||| |||||||| |||||||||||||||||| ||||||||||||
OLF: 6522 tcaagtcagtggcagagacggtcatcgccatgaggctgcccttctgcggccaccacgtga 6581

NOV1:579  tcagtcacttcacctgcaagatcctggcagtgctgaagctggcatgcggcaacacgtcgg 638
          |||| |||||||||||  |||||||||| |||||||||||| | || ||  |||| || |
OLF: 6582 tcagacacttcacctgtgagatcctggctgtgctgaagctgacctgtggtgacacctcag 6641

NOV1:639  tcagcgaagacttcctgctggcgggctccatcctgctgctgcctgtacccctggcattca 698
          ||||||| | |||||||||||  |||  ||||||| |||  |||||  |||||||| |  |||
OLF: 6642 tcagcgatgccttcctgctggtgggggccatcctcctgttgcctataccctgaccctca 6701

NOV1:699  tctgcctgtcctacttgctcatcctggccaccatcctgagggtgccctcggccgccaggt 758
          |||||||||||||| |||| |||||||||||||||||||||||||||||| ||| || ||
```

```
OLF: 6702 tctgcctgtcctacatgctgatcctggccaccatcctgagggtgccctcagccaccgggc 6761

NOV1:759  gctgcaaagccttctccacctgcttggcacacctggctgtagtgctgcttttctacggca 818
          || |||||||||||||||||||||| ||||||||||||||| || ||||||||||  |||
OLF: 6762 gcagcaaagccttctccacctgctcggcacacctggctgttgtcctgcttttctatagca 6821

NOV1:819  ccatcatcttcatgtacttgaagcccaagagtaaggaagcccacatctctgatgaggtct 878
          | ||||||||||||||| |||| |||||||| |||||||||  ||||| ||  ||||||
OLF: 6822 ctatcatcttcatgtacatgaaacccaagagcaaggaagcccggatctcagaccaggtct 6881

NOV1:879  tcacagtcctctatgccatggtcacgaccatgctgaaccccaccatctacagcctgagga 938
          | |||||||||||  ||  |||| ||  ||||||||||||||  ||||||||||||||||
OLF: 6882 ttacagtcctctacgctgtggtgacccccatgctgaaccccattatctacagcctgagga 6941

NOV1:939  acaaggaggtgaaggaggccgccaggaaggtgtggggcaggagtcgggcctccaggtgag 998
          |||||||||||||||| || |||||||| |  |||||||| ||  ||||||  |||||||
OLF: 6942 acaaggaggtgaaggaagcggccaggaaagcttggggcagcagatgggcctgtaggtgag 7001

NOV1:999  ggagggcggggctctg 1014 (SEQ ID No. 1)
          ||||||| ||||||||
OLF: 7002 ggagggcagggctctg 7017 (SEQ ID No. 15)
```

TABLE 4

```
NOV1: 1   MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY 60
          ||| ||| |||| |||||||||| ||||||| ||||||||||+|||+||  ||||+|
OLF: 1    MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGNTAIVAVSMLDARLHTPMY 60

NOV1: 61  FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY 120
          ||||||| |||||| |||||||||||||||||||  ||+|||||||||||||||||+ ||
OLF: 61   FFLGNLSILDICYTSTFVPLMLVHLLSSRKTISFTGCAVQMCLSLSTGSTECLLLAVMAY 120

NOV1:121  DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK 180
          |||||||||||| ||||||||++| ||+|||||  ||| | ||+|||||||||+ ||||+
OLF: 121  DRYLAICQPLRYPVLMSHRLCLMLAGASWVLCLFKSVAETVIAMRLPFCGHHVIRHFTCE 180

NOV1:181  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST 240
          ||||||| ||+||||+ ||| |+|||||+||  |||||+|||||||||||    |||||
OLF: 181  ILAVLKLTCGDTSVSDAFLLVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFST 240

NOV1:241  CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300
          | |||||||||| ||||||+||||||| |||+||||||+|| |||| ||||||||||||
OLF: 241  CSAHLAVVLLFYSTIIFMYMKPKSKEARISDQVFTVLYAVVTPMLNPIIYSLRNKEVKEA 300

NOV1:301  ARKVWGRSRASR 312 (SEQ ID No. 2)
          ||| ||   | |
OLF: 301  ARKAWGSRWACR 312 (SEQ ID No. 16)
```

Where '+' denotes similarity.

TABLE 5

```
macaca_OLF  ------------------------------------------------------------
NOV1        MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY
Mouse_OLF   MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGNTAIVAVSMLDARLHTPMY
Rat_OLF     ------------LLLGLSGYPKTEILYFVIVLVMYLVIHTGNGVLIIASIFDSHLHTPMY
Human_OLF   ---------MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMY
```

```
macaca_OLF      ------------------------------------------------------------
NOV1           FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY
Mouse_OLF       FFLGNLSILDICYTSTFVPLMLVHLLSSRKTISFTGCAVQMCLSLSTGSTECLLLAVMAY
Rat_OLF         FFLGNLSFLDICYTTSSVPSTLVSLISKKRNISFSGCTVQMFVGFAMGSTECLLLGMMAF
Human_OLF       FFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAF

macaca_OLF      ---PAICQPLRYRVLMNHRLCVLLVGAAWVLCLLKSVTETVIAMRLPFCGHHVVSHFTCE
NOV1            DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK
Mouse_OLF       DRYLAICQPLRYPVLMSHRLCLMLAGASWVLCLFKSVAETVIAMRLPFCGHHVIRHFTCE
Rat_OLF         DRYVAICNPLRYSVIMSKEVYVSMASASWFSGGINSVVQTSLAMRLPFCGNNVINHFTCE
Human_OLF       DRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCE
                    ***:**** *:*.:   : : .::*      **..  ::::*****.:*: ****:

macaca_OLF      ILAVLKLTCGNTSVSEVFLLVGSILLLPVPLAFICLSYLLILATILRVPSAAGCRKAFST
NOV1            ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST
Mouse_OLF       ILAVLKLTCGDTSVSDAFLLVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFST
Rat_OLF         VLAVLKLACADISLNIVTMVISNMAFLVLPLLLIFFSYVLILYTILRMNSASGRRKAFST
Human_OLF       ILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFST
                :******:*.: *:.   :    : :* :*: :* :**::*: ****: **    *.***

macaca_OLF      CSAHLAVVLLFYSTIIFTYMKPKSKE------AHISDEVFTVLYAMVTPML---------
NOV1            CLAHLAVVLLFYGTIIFMYLKPKSKE------AHISDEVFTVLYAMVTTMLNPTIYSLRN
Mouse_OLF       CSAHLAVVLLFYSTIIFMYMKPKSKE------ARISDQVFTVLYAVVTPMLNPIIYSLRN
Rat_OLF         CSAHLTVVVIFYGTIFSMYAKPKSQDLTGKDKFQTSDKIISLFYGVVTPMLNPIIYSLRN
Human_OLF       CSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLRN
                * ***:**::**.*::  * ****::          **::::.::*.:**.**

macaca_OLF      ------------------ (SEQ ID No. 17)
NOV1            KEVKEAARKVWGRSRASR (SEQ ID No. 2)
Mouse_OLF       KEVKEAARKAWGSRWACR (SEQ ID No. 18)
Rat_OLF         KDVKAAVKYILKQKYIP- (SEQ ID No. 19)
Human_OLF       KDVKAAVRRLLRPKGFTQ (SEQ ID No. 20)
```

Consensus key
* - single, fully conserved residue
: - conservation of strong groups
. - conservation of weak groups - no consensus

Because the OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade, NOV1 can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV1 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

NOV2

A NOV2 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. The NOV1 nucleic acid sequence (SEQ ID NO:1) was further analyzed by exon linking and the resulting sequence was identified as NOV2. A NOV2 nucleic acid and its encoded polypeptide includes the sequences shown in Table 6. The disclosed nucleic acid (SEQ ID NO: 3) is 1,050 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 59-61 and ends with a TGA stop codon at nucleotides 995-997. The representative ORF encodes a 312 amino acid polypeptide (SEQ ID NO: 4). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 3.

TABLE 6

CCCTGTACCCTCTCTCCTTCCATCCCAGCTGTGGACCATCTCTTCAGAACTCTGCAGC**ATG**GAGCCGCTCAACAGAA
CAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCA
GCCATGTACCTGGTGACCCTCCTGGGGAACACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCC
CGTGTACTTCTTCCTGGGCAACCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCC
ACCTCCTGTCATCCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCC
ACGGAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCACGTGCT
CATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGTCGGTGACTGAGATGG
TCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCCTGGCAGTGCTGAAG
CTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGCCTGTACCCCTGGC
ATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCT
TCTCCACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAG
AGTAAGGAAGCCCACATCTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCAT
CTACAGCCTGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGG**TGA**GGGA
GGGCGGGGCTCTGTACAGACGCAGGTCTCAGGTTAGTAGCTGAGGCCAT (SEQ ID NO. 3)

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLSTLDICYTPTF
VPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLL
KSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSA
ARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRS
RASR (SEQ ID NO. 4)

The target sequence previously identified, Accession Number AL135841 was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding seuqnce, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries. The resulting

amplicon was gel purified, clone, and sequenced to high redundancy to provide the sequence reported as NOV2.

The NOV2 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

The NOV2 nucleic acid sequence has homology (86% identity) with the mouse olfactory receptor gene cluster, OR17 and OR6 (OLF) (SEQ ID NO: 15)(GenBank Accession No:AJ251155), as shown in Table 7. Additionally, the NOV2 polypeptide has a high degree of homology (approximately 82% identity) to the mouse olfactory receptor 71 (OLF) (SEQ ID NO: 16) (GenBank Accession No:NP_062359), as shown in Table 8. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See Dryer and Berghard, 20 Trends in Pharmacological Sciences, 413 (1999).

OR proteins have seven transmembrane α-helices separated by three extracellular and three cytoplasmic loops, along with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence aligment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, NOV2 is predicted to have a seven transmembrane region and is similar in that region to representative olfactory receptor GPCRs of monkey (SEQ ID NO. 17) (GenBank Accession No:AAF40368), mouse (SEQ ID NO. 18) (GenBank Accession No:CAB55597), rat (SEQ ID NO:19) (GenBank Accession No: S29711), and human (SEQ ID NO. 20) (GenBank Accession No:CAB96728), as shown in Table 9.

TABLE 7

```
NOV2:99    tgaaaggattttctggctacccagccctggagcatctgctcttccctctgtgctcagcca 158
           |||| |||||||||||||||| ||||| ||||  || ||||| |||||||||||||  ||
OLF: 6102  tgaagggattttctggctacccggccctcgagcggctactctttcctctgtgctcagtca 6161

NOV2:159   tgtacctggtgaccctcctggggaacacagccatcatggcggtgagcgtgctagatatcc 218
           |||||||||||||  || |||||||||||||||||| ||||||||||| || | |||  ||
OLF: 6162  tgtacctggtgactctgctggggaacacagccatcgtggcggtgagcatgttggatgccc 6221

NOV2:219   acctgcacacgcccgtgtacttcttcctgggcaacctctctaccctggacatctgctaca 278
            |||||||||||||| ||||||| |||||||| ||||| || |   ||||||||||||||
OLF: 6222  gcctgcacacgcccatgtacttttcctgggtaacctttccattttggacatctgctaca 6281

NOV2:279   cgcccacctttgtgcctctgatgctggtccacctcctgtcatcccggaagaccatctcct 338
           |  | || ||||| || |||||||||||||||||||||||| |||||||||||||||||
OLF: 6282  catctacttttgtacccctgatgctggtccacctcctgtcgtcccggaagaccatctcct 6341

NOV2:339   ttgctgtctgtgccatccagatgtgtctgagcctgtccacgggctccacggagtgcctgc 398
           || | | ||||||| |||||||||||||||||||| |||||||||||||| |||||||||
OLF: 6342  ttacgggctgtgccgtccagatgtgtctgagcctctccacgggctccaccgagtgcctgc 6401

NOV2:399   tactggccatcacggcctatgaccgctacctggccatctgccagccactcaggtaccacg 458
```

```
          |  |||||  |||  ||||||||||||||||  |||||||  ||||||||||||||||||||  ||
OLF: 6402 tgttggccgtcatggcctatgaccgctacttggccatttgccagccactcaggtaccccg 6461

NOV2:459  tgctcatgagccaccggctctgcgtgctgctgatgggagctgcctgggtcctctgcctcc 518
          |||||||||||||| |||||||| || |||||   |||||  ||||||| |||||||||
OLF: 6462 tgctcatgagccacaggctctgcctgatgctggcaggagcctcctgggtgctctgcctct 6521

NOV2:519  tcaagtcggtgactgagatggtcatctccatgaggctgcccttctgtggccaccacgtgg 578
          ||||||| ||| | |||| |||||| ||||||||||||||||| ||||||||||||
OLF: 6522 tcaagtcagtggcagagacggtcatcgccatgaggctgcccttctgcggccaccacgtga 6581

NOV2:579  tcagtcacttcacctgcaagatcctggcagtgctgaagctggcatgcggcaacacgtcgg 638
          |||| ||||||||||  |||||||||| ||||||||||| | || || |||| || :
OLF: 6582 tcagacacttcacctgtgagatcctggctgtgctgaagctgacctgtggtgacacctcag 6641

NOV2:639  tcagcgaagacttcctgctggcgggctccatcctgctgctgcctgtacccctggcattca 698
          ||||||| | |||||||||||| |||  |||||||| ||| ||||| |||||||| |  |||
OLF: 6642 tcagcgatgccttcctgctggtgggggccatcctcctgttgcctatacccctgaccctca 6701

NOV2:699  tctgcctgtcctacttgctcatcctggccaccatcctgagggtgccctcggccgccaggt 758
          |||||||||||||| |||| |||||||||||||||||||||||||||||| ||| || ||
OLF: 6702 tctgcctgtcctacatgctgatcctggccaccatcctgagggtgccctcagccaccgggc 6761

NOV2: 59  gctgcaaagccttctccacctgcttggcacacctggctgtagtgctgcttttctacggca 818
          || ||||||||||||||||||||| |||||||||||||||| || |||||||||||  |||
OLF: 6762 gcagcaaagccttctccacctgctcggcacacctggctgttgtcctgcttttctatagca 6821

NOV2:819  ccatcatcttcatgtacttgaagcccaagagtaaggaagcccacatctctgatgaggtct 878
          | |||||||||||||||| |||| ||||||||| ||||||||||  |||||  ||  ||||||
OLF: 6822 ctatcatcttcatgtacatgaaacccaagagcaaggaagcccggatctcagaccaggtct 6881

NOV2:879  tcacagtcctctatgccatggtcacgaccatgctgaaccccaccatctacagcctgagga 938
          | |||||||||||| ||  |||| ||  |||||||||||||||  ||||||||||||||||
OLF: 6882 ttacagtcctctacgctgtggtgacccccatgctgaaccccattatctacagcctgagga 6941

NOV2:939  acaaggaggtgaaggaggccgccaggaaggtgtggggcaggagtcggcctccaggtgag 998
          |||||||||||||||| || ||||||||| |  |||||||| ||  ||||||  |||||||
OLF: 6942 acaaggaggtgaaggaagcggccaggaaagcttggggcagcagatgggcctgtaggtgag 7001

NOV2:999  ggagggcggggctctg 1014 (SEQ ID No. 3)
          ||||||| ||||||||
OLF: 7002 ggagggcagggctctg 7017 (SEQ ID No. 15)
```

TABLE 8

```
NOV2:1    MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY 60
          ||| ||| |||| ||||||||||| |||||||| ||||||||||||+|||+||  ||||+|
OLF: 1    MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGNTAIVAVSMLDARLHTPMY 60

NOV2:61   FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY 120
          ||||||| |||||| |||||||||||||||||||   ||+||||||||||||||||+ ||
OLF: 61   FFLGNLSILDICYTSTFVPLMLVHLLSSRKTISFTGCAVQMCLSLSTGSTECLLLAVMAY 120

NOV2:121  DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK 180
          |||||||||||| ||||||||++| ||+||||| ||| | ||+||||||||||+ ||||+
OLF: 121  DRYLAICQPLRYPVLMSHRLCLMLAGASWVLCLFKSVAETVIAMRLPFCGHHVIRHFTCE 180

NOV2:181  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST 240
          ||||||| ||+||||+ ||| |+|||||+|| |||||+|||||||||||||    |||||
OLF: 181  ILAVLKLTCGDTSVSDAFLLVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFST 240
```

```
NOV2:241 CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300
         | |||||||||| ||||||+|||||||  |||+|||||||+||  ||||  ||||||||||||
OLF: 241 CSAHLAVVLLFYSTIIFMYMKPKSKEARISDQVFTVLYAVVTPMLNPIIYSLRNKEVKEA 300

NOV2: 301 ARKVWGRSRASR 312 (SEQ ID No. 4)
          ||| ||   | |
OLF:  301 ARKAWGSRWACR 312 (SEQ ID No. 16)
```

Where '+' denotes similarity

TABLE 9

```
NOV2          MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY
macaca_OLF    ------------------------------------------------------------
Mouse_OLF     MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGNTAIVAVSMLDARLHTPMY
Rat_OLF       ------------LLLGLSGYPKTEILYFVIVLVMYLVIHTGNGVLIIASIFDSHLHTPMY
Human_OLF     ---------MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMY

NOV2          FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY
macaca_OLF    ------------------------------------------------------------
Mouse_OLF     FFLGNLSILDICYTSTFVPLMLVHLLSSRKTISFTGCAVQMCLSLSTGSTECLLLAVMAY
Rat_OLF       FFLGNLSFLDICYTTSSVPSTLVSLISKKRNISFSGCTVQMFVGFAMGSTECLLLGMMAF
Human_OLF     FFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAF

NOV2          DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK
macaca_OLF    ---PAICQPLRYRVLMNHRLCVLLVGAAWVLCLLKSVTETVIAMRLPFCGHHVVSHFTCE
Mouse_OLF     DRYLAICQPLRYPVLMSHRLCLMLAGASWVLCLFKSVAETVIAMRLPFCGHHVIRHFTCE
Rat_OLF       DRYVAICNPLRYSVIMSKEVYVSMASASWFSGGINSVVQTSLAMRLPFCGNNVINHFTCE
Human_OLF     DRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCE
                  ***:**** *:*.:   : : .::*      **..  ::::*****.:*: ****:

NOV2          ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST
macaca_OLF    ILAVLKLTCGNTSVSEVFLLVGSILLLPVPLAFICLSYLLILATILRVPSAAGCRKAFST
Mouse_OLF     ILAVLKLTCGDTSVSDAFLLVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFST
Rat_OLF       VLAVLKLACADISLNIVTMVISNMAFLVLPLLLIFFSYVLILYTILRMNSASGRRKAFST
Human_OLF     ILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFST
              :******:*.: *:.   :    : :* :*: :* :**::*: ****: **    *.***

NOV2          CLAHLAVVLLFYGTIIFMYLKPKSKE------AHISDEVFTVLYAMVTTMLNPTIYSLRN
macaca_OLF    CSAHLAVVLLFYSTIIFTYMKPKSKE------AHISDEVFTVLYAMVTPML---------
Mouse_OLF     CSAHLAVVLLFYSTIIFMYMKPKSKE------ARISDQVFTVLYAVVTPMLNPIIYSLRN
Rat_OLF       CSAHLTVVVIFYGTIFSMYAKPKSQDLTGKDKFQTSDKIISLFYGVVTPMLNPIIYSLRN
Human_OLF     CSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLRN
              * ***:**::**.*::  * ****::         **:::.::*.:**.**

NOV2          KEVKEAARKVWGRSRASR (SEQ ID No. 4)
macaca_OLF    ------------------ (SEQ ID No. 17)
Mouse_OLF     KEVKEAARKAWGSRWACR (SEQ ID No. 18)
Rat_OLF       KDVKAAVKYILKQKYIP- (SEQ ID No. 19)
Human_OLF     KDVKAAVRRLLRPKGFTQ (SEQ ID No. 20)
```

Consensus key

* - single, fully conserved residue

: - conservation of strong groups

. - conservation of weak groups - no consensus

The OR family of the GPCR superfamily is involved in the initial steps of the olfactory signal transduction cascade. Therefore, the NOV2 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on this relatedness to other known members of the OR family of the GPCR superfamily, NOV2 can be used to provide new diagnostic and/or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Moreover, nucleic acids, polypeptides, antibodies, and other compositions of the present invention are also useful in the treatment of a variety of diseases and pathologies, including but not limited to, those involving neurogenesis, cancer, and wound healing.

Hydrophobicity analysis confirms the prediction of the presence of seven transmembrane domains in NOV2. PSORT analysis predicts that NOV2 is localized to the plasma membrane. Likewise, SignalP analysis indicates that there is most likely a cleavage site between positions 47 and 48. Additionally, the following possible SNPs were identified:

82: T->G(11)
125218920(i), phred 40
125218923(i), phred 42
125219376(i), phred 40
125219632(i), phred 33
125219739(i), phred 33
125586244(i), phred 29
125586186(i), phred 34
125586110(i), phred 35
126544369(i), phred 45
125588716(i), phred 33
125219986(i), phred 37

91: C->T(11)
125218920(i), phred 37
125218923(i), phred 33
125219376(i), phred 37
125219632(i), phred 22
125219739(i), phred 37
125586244(i), phred 32
125586186(i), phred 25

125586110(i), phred 37

126544369(i), phred 37

125588716(i), phred 33

125219986(i), phred 37

150: C->G(10)

125218920(i), phred 45

125218923(i), phred 51

125219376(i), phred 38
125219632(i), phred 41
125219739(i), phred 51
125586244(i), phred 40
125586186(i), phred 45
125586110(i), phred 45
126544369(i), phred 40
125588716(i), phred 45

157: G->A(2)
125219739(i), phred 45
125586186(i), phred 45

246: G->C(10)
125218920(i), phred 40
125218923(i), phred 45
125219376(i), phred 42
125219632(i), phred 21
125219739(i), phred 45
125586244(i), phred 38
125586186(i), phred 32
125586110(i), phred 36
126544369(i), phred 45
125588716(i), phred 45

296: G->A(10)
125218920(i), phred 39
125218923(i), phred 36
125219376(i), phred 36
125219632(i), phred 36
125219739(i), phred 49
125586244(i), phred 36
125586186(i), phred 36
125586110(i), phred 39
126544369(i), phred 36
125588716(i), phred 36

406: A->G(2)
125586198(i), phred 38
125219755(i). phred 29

450: C->T(8)

125218920(i), phred 27
125218923(i), phred 24
125219376(i), phred 22
125219739(i), phred 29
125586244(i), phred 30
125586186(i), phred 19
126544369(i), phred 28
125530948(i), phred 27

562: A->G(5)
125531346(i), phred 29
125530963(i), phred 29
125531302(i), phred 49
125530948(i), phred 31
125531257(i), phred 24

662: C->T(6)
125531346(i), phred 36
125530963(i), phred 41
125531302(i), phred 37
125530948(i), phred 40
125531257(i), phred 40
126652213(i), phred 37

664: A->G(6)
125531346(i), phred 45
125530963(i), phred 41
125531302(i), phred 45
125530948(i), phred 45
125531257(i), phred 44
126652213(i), phred 45

667: A->T(6)
125531346(i), phred 37
125530963(i), phred 45
125531302(i), phred 45
125530948(i), phred 40
125531257(i), phred 45
126652213(i), phred 45

671: A->G(7)
125531283(i), phred 38
126652328(i), phred 45
126652243(i), phred 37
125531218(i), phred 45
125531233(i), phred 51
125531199(i), phred 45
125531268(i), phred 39

679: G->A(6)
125531346(i), phred 45
125530963(i), phred 45
125531302(i), phred 45
125530948(i), phred 45

125531257(i), phred 45
126652213(i), phred 37

776: C->T(6)
125531346(i), phred 41
125531302(i), phred 41
125530948(i), phred 45
126652243(i), phred 36
125531257(i), phred 45
126652213(i), phred 45
820: C->A(4)
125531346(i), phred 37
125530948(i), phred 40
125531257(i), phred 41
126652213(i), phred 45

NOV3

A NOV3 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV3 nucleic acid and its encoded polypeptide includes the sequences shown in Table 10. The disclosed nucleic acid (SEQ ID NO:5) is 1,031 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 22-24 and ends with a TGA stop codon at nucleotides 979-981. The representative ORF encodes a 319 amino acid polypeptide (SEQ ID NO:6). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO:5.

TABLE 10

TGATGGCAGAGGGGATATCAC**ATG**GAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGAGGCTCTC
TGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCTGCTGGGCAATGGGG
TCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCTTCCTAGGGAACCTCTCCTTCCTG
GACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGCTTTTTGACTCCCCAGGAAACCATCTCCTTCTC
AGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCATGGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTG
ATCGCTATGTGGCCATCTGCAACCCCCTTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCC
AGCTCCTGGGCTATTGGTGGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAA
TGTCATCAACCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTCATCATCACC
ACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGCCCACCTCACCGTGGTGAT
CGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGGACTCCATGGGAGCAGACAAAGAGGATC
TTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACCCCGATGCTCAACCCCATCATCTATAGCCTGAGGAAC
AAGGATGTGAAGGCTGCTGTGAGGAGACTGCTGAGACCAAAAGGCTTCACTCAG**TGA**TGGTGGAAGGGTCCTCTGTG
ATTGTCACCCACATGGAAGTAAGGAATCAC (SEQ ID NO:5)

TABLE 11. Polypeptide Sequence Encoded by Nucleic Acid Sequence of Table 10.

MEKANETSPVMGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLSFLDICFTTS
SVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVIMSKAAYMPMAASSWAIGG
AASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPS
AEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAV
RRLLRPKGFTQ (SEQ ID NO:6)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV3 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue. A NOV3 nucleic acid was identified on human chromosome 1.

The NOV3 nucleic acid sequence is homologous to (100 % identity) to a human genomic clone corresponding to chromosome 9p13.1- 13.3 (CHR9) (SEQ ID NO: 21) (GenBank Accession No:AL135841), as is shown in Table 12. Also, the NOV3 polypeptide has homology (approximately 88 % identity) to the human olfactory receptor, family 2, subfamily S, member 2 (OLF) (SEQ ID NO: 20) (GenBank Accession No:CAB96728), as is shown in Table 13. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. (Dryer and Berghard, 20 Trends in Pharmacological Sciences 413 (1999)). OR proteins have seven transmembrane α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence aligment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, NOV3 is predicted to have a seven transmembrane region and is similar in that region to representative olfactory receptor GPCRs of human (SEQ ID NO. 20) (GenBank Accession No:CAB96728), rat (SEQ ID NO. 22) (GenBank Accession No:AAC64588), and mouse (SEQ ID NO. 23) (GenBank Accession No:CAB96147), as shown in Table 14.

TABLE 12

```
NOV3:     1 tgatggcagaggggatatcacatggaaaaagccaatgagacctcccctgtgatggggttc 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82721 tgatggcagaggggatatcacatggaaaaagccaatgagacctcccctgtgatggggttc 82662

NOV3:    61 gttctcctgaggctctctgcccacccagagctggaaaagacattcttcgtgctcatcctg 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82661 gttctcctgaggctctctgcccacccagagctggaaaagacattcttcgtgctcatcctg 82602

NOV3:   121 ctgatgtacctcgtgatcctgctgggcaatggggtcctcatcctggtgaccatccttgac 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82601 ctgatgtacctcgtgatcctgctgggcaatggggtcctcatcctggtgaccatccttgac 82542

NOV3:   181 tcccgcctgcacacgcccatgtacttcttcctagggaacctctccttcctggacatctgc 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82541 tcccgcctgcacacgcccatgtacttcttcctagggaacctctccttcctggacatctgc 82482

NOV3:   241 ttcactacctcctcagtcccactggtcctggacagctttttgactccccaggaaaccatc 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82481 ttcactacctcctcagtcccactggtcctggacagctttttgactccccaggaaaccatc 82422
```

```
NOV3: 301   tccttctcagcctgtgctgtgcagatggcactctcctttgccatggcaggaacagagtgc 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82421 tccttctcagcctgtgctgtgcagatggcactctcctttgccatggcaggaacagagtgc 82362

NOV3: 361   ttgctcctgagcatgatggcatttgatcgctatgtggccatctgcaacccccttaggtac 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82361 ttgctcctgagcatgatggcatttgatcgctatgtggccatctgcaacccccttaggtac 82302

NOV3: 421   tccgtgatcatgagcaaggctgcctacatgcccatggctgccagctcctgggctattggt 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9  82301 tccgtgatcatgagcaaggctgcctacatgcccatggctgccagctcctgggctattggt 82242

NOV3: 481   ggtgctgcttccgtggtacacacatccttggcaattcagctgcccttctgtggagacaat 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9. 82241 ggtgctgcttccgtggtacacacatccttggcaattcagctgcccttctgtggagacaat 82182

NOV3: 541   gtcatcaaccacttcacctgtgagattctggctgttctaaagttggcctgtgctgacatt 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82181 gtcatcaaccacttcacctgtgagattctggctgttctaaagttggcctgtgctgacatt 82122

NOV3· 601   tccatcaatgtgatcagcatggaggtgacgaatgtgatcttcctaggagtcccggttctg 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82121 tccatcaatgtgatcagcatggaggtgacgaatgtgatcttcctaggagtcccggttctg 82062

NOV3  661   ttcatctctttctcctatgtcttcatcatcaccaccatcctgaggatcccctcagctgag 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82061 ttcatctctttctcctatgtcttcatcatcaccaccatcctgaggatcccctcagctgag 82002

NOV3· 721   gggaggaaaaaggtcttctccacctgctctgcccacctcaccgtggtgatcgtcttctac 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 82001 gggaggaaaaaggtcttctccacctgctctgcccacctcaccgtggtgatcgtcttctac 81942

NOV3  781   gggaccttattcttcatgtatgggaagcctaagtctaaggactccatgggagcagacaaa 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 81941 gggaccttattcttcatgtatgggaagcctaagtctaaggactccatgggagcagacaaa 81882

NOV3: 841   gaggatctttcagacaaactcatcccccttttctatggggtggtgaccccgatgctcaac 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 81881 gaggatctttcagacaaactcatcccccttttctatggggtggtgaccccgatgctcaac 81822

NOV3· 901   cccatcatctatagcctgaggaacaaggatgtgaaggctgctgtgaggagactgctgaga 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 81821 cccatcatctatagcctgaggaacaaggatgtgaaggctgctgtgaggagactgctgaga 81762

NOV3. 961   ccaaaaggcttcactcagtgatggtggaagggtcctctgtgattgtcacccacatggaag 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHR9: 81761 ccaaaaggcttcactcagtgatggtggaagggtcctctgtgattgtcacccacatggaag 81702

NOV3: 1021  taaggaatcac 1031 (SEQ ID NO. 5)
            ||||||||||||
CHR9: 81701 taaggaatcac 81691 (SEQ ID NO. 21)
```

TABLE 13

```
NOV3:11  MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLSFL 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OLF: 1   MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLSFL 60

NOV3:71  DICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNP 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OLF: 61  DICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNP 120

NOV3:131 LRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLAC 190
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
OLF: 121 LRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLAC 180

NOV3:191 ADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVI 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OLF: 181 ADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVI 240

NOV3:251 VFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRR 310
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OLF: 241 VFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRR 300

NOV3:311 LLRPKGFTQ 319 (SEQ ID NO. 6)
         |||||||||
OLF: 301 LLRPKGFTQ 309 (SEQ ID NO. 20)
```

Table 14

```
NOV3          MEKANETSPVMGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPM
Human_OLF     ----------MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPM
rat_OLF       ------------------------------------------------------------
mouse_OLF     MDRSNETAPLSGFILLGLSAHPKLEKTFFVLILMMYLVILLGNGVLILVSILDSHLHTPM

NOV3          YFFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMA
Human_OLF     YFFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMA
rat_OLF       ----SNLSFLDICYTTSSVPLILGSFLTPRKTISFSGCAVQMFLSFAMGATECVLLSMMA
mouse_OLF     YFFLGNLSFLDICYTTSSVPLILDSFLTPRKTISFSGCAVQMFLSFAMGATECVLLSMMA
                  .********:*******:*.*****::***** ***** *****..***:******

NOV3          FDRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTC
Human_OLF     FDRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTC
rat_OLF       FDRYVAICNPLRYPVVMSKAVYVPMATGSWAAGIAASLVQTSLAMRLPFCGDNVINHFTC
mouse_OLF     FDRYVAICNPLRYPVVMNKAAYVPMAASSWAGGITNSVVQTSLAMRLPFCGDNVINHFTC
              **************.*:*.**.*:***:.*** * : *:*:****::**************

NOV3          EILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFS
Human_OLF     EILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFS
rat_OLF       EILAVLKLACADISINIISMGVTNVIFLGVPVLFISFSYIFILSTILRIPSAEGRKKAFS
mouse_OLF     EILAVLKLACADISINVISMVVANMIFLAVPVLFIFVSYVFILVTILRIPSAEGRKKAFS
              *****************:*** *:*:***.******  **:**: *************.**

NOV3          TCSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLR
Human_OLF     TCSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLR
rat_OLF       TCSAHLTVVIVFYGTILFMYGKPKSKDPLGADKQDPADKLISLFYGVLTPM---------
mouse_OLF     TCSAHLTVVLVFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGVVTPMLNPIIYSLR
              *********:*****::**********.:****:* :****.*****:***

NOV3          NKDVKAAVRRLLRPKGFTQ (SEQ ID NO. 6)
Human_OLF     NKDVKAAVRRLLRPKGFTQ (SEQ ID NO. 20)
rat_OLF       ------------------- (SEQ ID NO. 22)
mouse_OLF     NKDVRAAVRNLVGQKHLTE (SEQ ID NO. 23)
```

Consensus key

* - single, fully conserved residue
: - conservation of strong groups
. - conservation of weak groups
  - no consensus

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV3 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV3 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in treating and/or diagnosing a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

Hydrophobicity analysis confirms the prediction of the presence of seven transmembrane domains in NOV3. PSORT analysis predicts that NOV3 is likely localized in the plasma membrane, the Golgi body, the endoplasmic reticulum (membrane), and the endoplasmic reticulum (lumen). Likewise, SignalP analysis indicates that there is most likely a cleavage site between positions 44 and 45.

NOV1, NOV2 and NOV3 are also described in USSN 09/777,789, filed February 6, 2001, hereby incorporated by reference in its entirety.

NOV4

A NOV4 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV4 nucleic acid and its encoded polypeptide include the sequence shown in Table 15. The disclosed nucleic acid (SEQ ID NO: 7) is 1,050 nucleotides in length.

TABLE 15. Nucleotide sequence encoding the Olfactory Receptor-like protein of the invention.

CCCTGTACCCTCTCTCCTTCCATCCCAGCTGTGGACCATCTCTTCAGAACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGT
CCGAGTTCTTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACC
CTCCTGGGGAACACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCCGTGTACTTCTTCCTGGGCAACCTCTC
TACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCATCCCGGAAGACCATCTCCTTTGCTG
TCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACGGAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTG

GCCATCTGCCAGCCACTCAGGTACCACGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCT
CCTCAAGTCGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCC
TGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGCCTGTACCC
CTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTC
CACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCC
ACATCTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCCTGAGGAACAAGGAG
GTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGGAGGGCGGGGCTCTGTACAGACGCAGGTCTCAG
GTTAGTAGCTGAGGCCAT (SEQ ID NO: 7)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV4 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV4 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in treating and/or diagnosing a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

TABLE 16. Protein sequence encoded by the NOV4 coding sequence of Table 15.

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLL
SSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHH
VVSHFTCKILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFM
YLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR (SEQ ID NO: 8)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV4 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue. A NOV4 nucleic acid was identified on human chromosome 1. Table 17 depicts the association of the variant sequence described in Table 16 with a phenotypic trait.

TABLE 17. Association of variant sequence described in Tables 16 with a phenotypic trait.

Variant	Associated Trait	p value (signficiance)	Shift in trait per allele
13019736	serum gamma-glutamyl transpeptidase	0.0001	- 0.4 s.d.
13019736	bone density	0.0005	+ 0.4 s.d.
13019736	serum calcium	0.002	- 0.4 s.d.

The NOV4 polypeptide has homology to the human olfactory receptor, family 2, subfamily S, member 2 (OLF) (GenBank Accession No:CAB96728)(SEQ ID NO:20), as shown by its relationship to NOV1-NOV3 polypeptide sequences. Further ClustalW analyses of the NOVX sequences are shown in Example 5.

Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. (Dryer and Berghard, 20 Trends in Pharmacological Sciences 413 (1999)). OR proteins have seven transmembrane α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence aligment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, NOV4 is predicted to have a seven transmembrane region and is similar in that region to representative olfactory receptor GPCRs of human (SEQ ID NO. 20) (GenBank Accession No:CAB96728), rat (SEQ ID NO. 22) (GenBank Accession No:AAC64588), and mouse (SEQ ID NO. 23) (GenBank Accession No:CAB96147), as shown in Table 14 above.

The nucleic acid encoding the NOV4 protein, or fragments thereof, may be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "NOVX Antibodies" section below. The disclosed NOV4 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV4 epitope is from about amino acids 1 to 20. In another embodiment, a NOV4 epitope is from about amino acids 125 to 140. In further embodiments, NOV4 epitopes are from about amino acids 250 to 270 and from about amino acids 275 to 312.

It is noted that NOV4, which utilizes new internal identification number CG54212-01, is identical in its nucleotide and amino acid sequences to NOV1 and NOV2, which utilize the former internal identification number, AL135841_B.

NOV5

A NOV5 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV5 nucleic acid and its encoded polypeptide include the sequence shown in Tables 18 and 19. The disclosed nucleic acid (SEQ ID NO: 9) is 1,050 nucleotides in length, and is a single nucleotide polymorphism variant of SEQ ID NO: 7 at position 236 where C replaces T (see underlined nucleotide).

TABLE 18. Variant of nucleotide sequence of Table 15 (nucleotide sequence of variant 13019736, underlined, T/C).

```
CCCTGTACCCTCTCTCCTTCCATCCCAGCTGTGGACCATCTCTTCAGAACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGT
CCGAGTTCTTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACC
CTCCTGGGGAACACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCCGTGCACTTCTTCCTGGGCAACCTCTC
TACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCATCCCGGAAGACCATCTCCTTTGCTG
TCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACGGAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTG
GCCATCTGCCAGCCACTCAGGTACCACGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCT
CCTCAAGTCGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCC
TGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGCCTGTACCC
CTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTC
CACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCC
ACATCTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCCTGAGGAACAAGGAG
GTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGGAGGGCGGGGCTCTGTACAGACGCAGGTCTCAG
GTTAGTAGCTGAGGCCAT (SEQ ID NO: 9)
```

TABLE 19. Polypeptide Sequence Encoded by Variant Nucleic Acid Sequence of Table 18 (Alteration effect: Tyr to His).

```
MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVHFFLGNLSTLDICYTPTFVPLMLVHLL
SSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHH
VVSHFTCKILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFM
YLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR (SEQ ID NO 10)
```

SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; window size (number of bases in a view) is 10; allowed number of mismatches in a window is 2; minimum SNP base quality (PHRED score) is 23; minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected

and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

A method for confirming novel SNPs comprises employing a validated method know as "pyrosequencing". Detailed protocols for pyrosequencing can be found in Alderborn et al. (Alderborn et al., "Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing,"10(8) Genome Research 1249-1265 (2000)).

NOV6

A NOV6 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV6 nucleic acid and its encoded polypeptide include the sequences shown in Tables 20 and 21. The disclosed nucleic acid (SEQ ID NO: 11) is 1,008 nucleotides in length.

TABLE 20. Nucleotide sequence encoding an Olfactory Receptor-like NOV6 protein.

```
AGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTCACCTTGGGAAGTGGCTTC
ATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGCTACAATTACAATCCTATACTTGTTGGCCCTGATCAG
CAATGGCCTACTGCTCCTGGCTATCACCATGGAAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGG
ACCTCCTGTTCACATCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTGTGCC
CTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCTATGACAGGTATGTGGCCATTTG
TCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTGGCTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTG
CCCTAATATATACCGTGTATACCATGCACTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTG
CTGAAGTTGGCCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCTCTTGCTGC
TATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGGGAGGAAGAAAGCCCTTGTCACCTGCT
CTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCACATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAA
GACAACATCATCTCTGTTTTCTACACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCG
GGCCTTGAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA (SEQ ID NO: 11)
```

TABLE 21. Protein sequence encoded by the NOV6 coding sequence of Table 20.

```
MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITLYLLALISNGLLLLAITMEARLHMPMYLLLGQLSLMDLLFTSVVT
PKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTM
HYPFCRAQEIRHLLCEIPHLLKLACADTSRYELMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGM
FYGAATFMYVLPSSFHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL (SEQ ID NO: 12)
```

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV6 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in treating and/or diagnosing a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

Table 22 depicts the association of the variant sequence of Table 21 with a phenotypic trait.

Table 22. Association of variant NOV6 sequence with a phenotypic trait.

Variant	Associated Trait	p value (signficiance)	Shift in trait per allele
13373788	serum Apolipoprotein(a)	0.0001	+ 0.4 sd

NOV7

A NOV7 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV7 nucleic acid and its encoded polypeptide include the sequence shown in Tables 23 and 24. The disclosed nucleic acid (SEQ ID NO: 13) is 1,008 nucleotides in length, and is a single nucleotide polymorphism variant of SEQ ID NO: 11 at position 278 where C replaces the T found in SEQ ID NO: 11.

TABLE 23. Variant of nucleotide sequence of Table 20: nucleotide sequence variant 13373788 (underlined), T/C.

AGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTCACCTTGGGAAGTGGCTTC
ATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGCTACAATTACAATCCTATACTTGTTGGCCCTGATCAG
CAATGGCCTACTGCTCCTGGCTATCACCATGGAAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGG
ACCTCCTGTTCACATCTGT**C**GTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTGTGCC
CTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCTATGACAGGTATGTGGCCATTTG
TCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTGGCTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTG
CCCTAATATATACCGTGTATACCATGCACTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTG
CTGAAGTTGGCCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCTCTTGCTGC
TATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGGGAGGAAGAAAGCCCTTGTCACCTGCT
CTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCACATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAA
GACAACATCATCTCTGTTTTCTACACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCG
GGCCTTGAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA (SEQ ID NO: 13)

TABLE 24. Polypeptide sequence encoded by variant nucleic acid sequence (underlined) of Table 23.

MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMYLLLGQLSLMDLLFTS**V**VT
PKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTM
HYPFCRAQEIRHLLCEIPHLLKLACADTSRYELMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGM
FYGAATFMYVLPSSFHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL (SEQ ID NO: 14)

Nucleotide change is silent, with no coding change in resultant NOV7 polypeptide.

SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; window size (number of

bases in a view) is 10; allowed number of mismatches in a window is 2; minimum SNP base quality (PHRED score) is 23; minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

A method for confirming novel SNPs comprises employing a validated method know as "pyrosequencing". Detailed protocols for pyrosequencing can be found in Alderborn et al. (Alderborn et al., "Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing,"10(8) Genome Research 1249-1265 (2000)).

NOVX Nucleic Acids

The nucleic acids of the invention include those that encode a NOVX polypeptide or protein. As used herein, the terms polypeptide and protein are interchangeable.
In some embodiments, a NOVX nucleic acid encodes a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein described herein relates to the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a

step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

Among the NOVX nucleic acids is the nucleic acid whose sequence is provided in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or a fragment thereof, any of whose bases may be changed from the corresponding bases shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, while still encoding a protein that maintains at least one of its NOVX-like activities and physiological functions (*i.e.,* modulating angiogenesis, neuronal development). The invention further includes the complement of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX proteins or biologically active portions thereof. Also included are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a

vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, as a hybridization probe, NOVX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule

less than 100 nt in length would further comprise at lease 6 contiguous nucleotides of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, or 13, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotide units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of NOVX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY (1993), and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Smith and Waterman, 2 Adv. Appl. Math.482-489 (1981), which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a NOVX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g* , mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, as well as a polypeptide having NOVX activity. Biological activities of the NOVX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human NOVX polypeptide.

The nucleotide sequence determined from the cloning of the human NOVX gene allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX

homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or of a naturally occurring mutant of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12.

Probes based on the human NOVX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

A "polypeptide having a biologically active portion of NOVX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 that encodes a polypeptide having a NOVX biological activity (biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX. For example, a nucleic acid fragment encoding a biologically active portion of NOVX can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of NOVX includes one or more regions.

NOVX Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, due to the degeneracy of the genetic code. These nucleic acids thus encode the same NOVX protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, *e.g.*, the polypeptide of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14. In another embodiment, an isolated nucleic acid molecule of the

invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14.

In addition to the human NOVX nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of NOVX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a NOVX protein, preferably a mammalian NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in NOVX that are the result of natural allelic variation and that do not alter the functional activity of NOVX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human NOVX cDNA can be isolated based on its homology to human membrane-bound NOVX. Likewise, a membrane-bound human NOVX cDNA can be isolated based on its homology to soluble human NOVX.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high

stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y.,6.3.1-6.3.6 (1989). Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X

SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. (See, *e.g.*, Ausubel et al., Eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, (1993), and Kriegler, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY (1990).

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, 26, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art, for example, as employed for cross-species hybridizations. (See, *e.g.*, Ausubel et al., Eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, (1993) and Kriegler, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY (1990); Shilo and Weinberg, 78 Proc Natl Acad Sci USA 6789-6792 (1993).

Conservative mutations

In addition to naturally-occurring allelic variants of the NOVX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, thereby leading to changes in the amino acid sequence of the encoded NOVX protein, without altering the functional ability of the NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of NOVX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, yet

retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in NOVX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant NOVX protein can be assayed for (1) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant NOVX protein and a NOVX receptor; (3) the ability of a mutant NOVX protein to bind to an intracellular target

protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind NOVX protein; or (5) the ability to specifically bind an anti-NOVX protein antibody.

Antisense NOVX Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14 or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of human NOVX corresponds to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding NOVX disclosed herein (*e.g.*, SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or

variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein, *e g* , by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules,

vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al., 15 Nucleic Acids Res 6625-6641 (1987)). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., 15 Nucleic Acids Res 6131-6148(1987)) or a chimeric RNA -DNA analogue (Inoue et al., 215 FEBS Lett 327-330(1987)).

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

NOVX Ribozymes and PNA moieties

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes such as hammerhead ribozymes can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA (Haselhoff and Gerlach, 334 Nature 585-591 (1988)). A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX DNA disclosed herein (*i.e.*, SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. (See U.S. Pat. No. 4,987,071 and U.S. Pat. No. 5,116,742, both to Cech et al.). Alternatively, NOVX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. (Bartel et al., 261 Science 1411-1418 (1993).

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target

cells. (Helene, 6 Anticancer Drug Des. 569-84 (1991); Helene et al., 660 Ann. N.Y. Acad. Sci. 27-36 (1992); and Maher, 14 Bioassays 14: 807-15 (1992)).

In various embodiments, the nucleic acids of NOVX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (Hyrup et al., 4 Bioorg. Med. Chem. 5-23 (1996)). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (Id.) and Perry-O'Keefe et al., 93 PNAS 14670-675 (1996).

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup et al., *ibid.*); or as probes or primers for DNA sequence and hybridization (Hyrup et al., Id.; Perry-O'Keefe, *ibid.*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup et al., Id.). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup et al. and Finn *et al.* (Hyrup et al., Id.; Finn et al., 24 Nucl. Acids Res. 3357-63 (1996)). For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e g.*, 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA

and the 5' end of DNA (Mag et al., 17 Nucl Acid Res 5973-88(1989)). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., *ibid.*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. (See, Petersen et al., 5 Bioorg. Med. Chem. Lett. 1119-11124(1975)).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger et al., 86 Proc. Natl. Acad. Sci. U.S.A.6553-6556(1989); Lemaitre et al., 84 Proc. Natl. Acad. Sci. 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol et al., 6 BioTechniques 958-976(1988)) or intercalating agents. (See, *e.g.*, Zon, 5 Pharm. Res. 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

NOVX Polypeptides

A NOVX polypeptide of the invention includes the NOVX-like protein whose sequence is provided in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14 while still encoding a protein that maintains its NOVX-like activities and physiological functions, or a functional fragment thereof. In some embodiments, up to 20% or more of the residues may be so changed in the mutant or variant protein. In some embodiments, the NOVX polypeptide according to the invention is a mature polypeptide.

In general, a NOVX -like variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are

polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX protein having less than about 30% (by dry weight) of non-NOVX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX protein, still more preferably less than about 10% of non-NOVX protein, and most preferably less than about 5% non-NOVX protein. When the NOVX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically active portions of a NOVX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the NOVX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14 that include fewer amino acids than the full length NOVX proteins, and exhibit at least one activity of a

NOVX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically active portion of a NOVX protein can be a polypeptide that is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a NOVX protein of the present invention may contain at least one of the above-identified domains conserved between the NOVX proteins, *e.g.* TSR modules. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, and retains the functional activity of the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, and retains the functional activity of the NOVX proteins of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14.

SNPs

SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2; Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

A method for confirming novel SNPs comprises employing a validated method know as "pyrosequencing". Detailed protocols for pyrosequencing can be found in Alderborn et al. (Alderborn et al., "Determination of Single Nucleotide Polymorphisms by Real-time

Pyrophosphate DNA Sequencing,"10(8) Genome Research 1249-1265 (2000)).

In brief, pyrosequencing is a real time primer extension process of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase-mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes proportional to the number of added bases, up to about four bases. To allow processivity of the method dNTP excess is degraded by apyrase, which is also present in the starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels.

The association of a SNP with a defined phenotypic trait is discovered through statistical genetic analysis of the SNP in a population sample of humans in which the phenotypic trait under investigation has been characterized. Such a population may consist of unrelated individuals, or of related individuals such as sibling pairs (including dizygotic or monozygotic twins), offspring & parents, or other familial structures comprised of genetically related individuals. These populations may be ascertained based upon the presence of one or more disease-affected individual(s) within each family, or may be ascertained as an epidemiologic sample representing the entire population. The phenotypic traits may be any observable or measurable characteristic
of humans, including but not limited to biochemical assays, assays of physiological function or performance, and clinical measures of growth and development such as body mass index. Specific analytic methods used depend upon the specific family structures, such as QTDT for sibling pairs (Abecasis et al., "A General Test of Association for Quantitative Traits in Nuclear Families," 66 Am. J. Hum. Genet. 279-292 (2000)).

Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid

residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. (Needleman and Wunsch, 48 J. Mol. Biol. 443-453 (1970)). Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Chimeric and fusion proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to NOVX, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically active portions of a NOVX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame to each other. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

For example, in one embodiment a NOVX fusion protein comprises a NOVX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate NOVX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (*i.e.*, glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX.

In another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated NOVX ligand of the invention is the NOVX receptor. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e,g.,* cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival, as well as acute and chronic inflammatory disorders and hyperplastic wound healing, *e.g.* hypertrophic scars and keloids.

Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al., Eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY (1992)). Moreover, many expression vectors are commercially available that already encode a fusion moiety, for example, a GST polypeptide. A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX agonists and antagonists

The present invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the NOVX protein. An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade that includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX protein that function as either NOVX agonists (mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the NOVX protein for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods that can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see Narang, 39 Tetrahedron 3 (1983); Itakura et al., 53 Annual Rev. Biochem. 323 (1984); Itakura et al., 198 Science 1056 (1984); Ike et al., 11 Nucl. Acid Res. 477 (1983).

Polypeptide libraries

In addition, libraries of fragments of the NOVX protein coding sequence can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely

used techniques, which are amenable to high throughput analysis. for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants (Arkin and Yourvan, 89 PNAS 7811-7815 (1992); Delgrave et al., 6 Protein Engineering 327-331 (1993)).

NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e* , molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30

amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. (See Hopp and Woods, 78 Proc. Nat. Acad. Sci. USA 3824-3828 (1991); Kyte and Doolittle, 157 J. Mol. Biol. 105-142 (1982), each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E., and Lane D., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1988) incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not

limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (D. Wilkinson, 14(8) The Scientist 25-28 (2000)).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein (Kohler and Milstein, 256 Nature 495 (1975)). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized *in vitro*.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal

Antibodies: Principles and Practice, Academic Press, pp. 59-103 (1986)). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells. Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, 133 J. Immunol. 3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, pp. 51-63(1987)).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard (Munson and Pollard, 107 Anal. Biochem. 220 (1980)). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown *iv vivo* as ascites in a mammal.
The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for

example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, 368 Nature 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., 321 Nature 522-525 (1986); Riechmann et al., 332 Nature 323-327 (1988); Verhoeyen et al., 239 Science 1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither

in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Id. and Presta, 2 Curr. Op. Struct. Biol. 593-596 (1992)).

In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Id. and Presta, 2 Curr. Op. Struct. Biol. 593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor et al., 4 Immunol. Today 72 (1983)) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96 (1985)). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (Cote et al., 80 Proc Natl Acad Sci USA 2026-2030 (1983)) or by transforming human B-cells with Epstein Barr Virus *in vitro* (Cole, et al., In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96 (1985)).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the

endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. 10 Bio/Technology 779-783 (1992); Lonberg et al., 368 Nature 856-859 (1994); Morrison, 368 Nature 812-13 (1994); Fishwild et al., 14 Nature Biotechnology 845-51 (1996); Neuberger 14 Nature Biotechnology 826 (1996); and Lonberg and Huszar 13 Intern. Rev. Immunol. 65-93 (1995).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker;

and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (Huse et al., 246 Science 1275-1281(1989)) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, 305 Nature 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a

potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.* (Traunecker et al., 10 EMBO J. 3655-3659 (1991)).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. (For further details of generating bispecific antibodies see, for example, Suresh et al., 121 Methods in Enzymology 210 (1986)).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g* tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e g.* $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al. describes a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments (Brennan et al., 229 Science 81 (1985)). These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific

antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al. describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule (Shalaby et al., 175 J. Exp. Med. 217-225 (1992)). Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers (Kostelny et al., 148(5) J. Immunol. 1547-1553 (1992)). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al. has provided an alternative mechanism for making bispecific antibody fragments (Hollinger et al., 90 Proc. Natl. Acad. Sci. USA 6444-6448 (1993)). The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. (See Gruber et al., 152 J. Immunol. 5368 (1994)).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. (Tutt et al., 147 J. Immunol. 60 (1991)).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies

possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC) (Caron et al., 176 J. Exp Med.1191-1195 (1992) and Shopes, J., 148 Immunol. 2918-2922 (1992)). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. (Wolff et al., 53 Cancer Research 2560-2565 (1993)). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. (Stevenson et al., 3 Anti-Cancer Drug Design 219-230 (1989)).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al. (Vitetta et al., 238 Science 1098 (1987)). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be

ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g* , tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech, Inc.) (Smith and Johnson, 67 Gene 31-40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., 69 Gene 301-315 (1988)) and pET 11d (Studier et al., 185 GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY, Academic Press, San Diego, Calif. 60-89 (1990)).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. (See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. 119-128 (1990)). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada, et al., 20 Nucl. Acids Res. 2111-2118 (1992)). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, et al., 6 EMBO J. 229-234 (1987)), pMFa (Kurjan and Herskowitz, 30 Cell 933-943 (1982)), pJRY88 (Schultz et al., 54 Gene 54: 113-123 (1987)), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, et al., 3 Mol. Cell. Biol. 2156-2165(1983)) and the pVL series (Lucklow and Summers, 170 Virology 31-39 (1989)).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 329 Nature 840(1987)) and pMT2PC (Kaufman, et al., 6 EMBO J. 187-195 (1987)). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1 Genes Dev. 268-277 (1987)), lymphoid-specific promoters (Calame and Eaton, 43 Adv. Immunol. 235-275 (1988)), in particular promoters of T cell receptors (Winoto and Baltimore, 8 EMBO J. 729-733 (1989)) and immunoglobulins (Banerji, et al., 33 Cell 729-740 (1983); Queen and Baltimore, 33 Cell 741-748 (1983)), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 86 Proc. Natl. Acad. Sci. USA 5473-5477 (1989)), pancreas-specific promoters (Edlund, et al., 230 Science 912-916(1985)), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 249 Science 374-379 (1990)) and the α-fetoprotein promoter (Campes and Tilghman, 3 Genes Dev. 537-546 (1989)).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," 1(1) Reviews-Trends in Genetics (1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human, Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor

Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that

remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. Sequences including SEQ ID NO: 1, 3, 5, 7, 9, 11 or 12 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191 (see also Hogan, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1986)). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the DNA of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one

embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. See Thomas, et al. for a description of homologous recombination vectors (Thomas, et al., 51 Cell 503 (1987)). The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. (*See, e.g.*, Li, et al., 69 Cell 915 (1992)).

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. (*See, e.g.*, Bradley In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH. Robertson, ed. IRL, Oxford, pp. 113-152 (1987)). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley. (Bradley, 2 Curr. Opin. Biotechnol. 823-829 (1991); PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *see, e.g.*, Lakso, *et al.* (Lakso et al., 89 Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. (O'Gorman, et al., 251 Science 1351-1355 (1991)). If a cre/loxP

recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.* (Wilmut et al., 385 Nature 810-813 (1997)). In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., 82 Proc. Natl. Acad. Sci. USA 3688 (1985); Hwang et al., 77 Proc. Natl Acad. Sci. USA 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., 257 J. Biol. Chem., 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., 81(19) J. National Cancer Inst. 1484 (1989).

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*. inhalation), transdermal (*i.e.,* topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of

microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include. for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (see, e.g., Chen, et al., 91 Proc. Natl. Acad. Sci. USA 3054-3057(1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington, The Science And Practice Of Pharmacy, 19th ed. (Alfonso R. Gennaro, et al., Editors) Mack Pub. Co., Easton, Pa. (1995); Drug Absorption Enhancement: Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa. (1994); and Peptide And Protein Drug Delivery in 4 Advances In Parenteral Sciences , M. Dekker, New York (1991). If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. (See Marasco et al., 90 Proc. Natl. Acad. Sci. USA 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended. The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules,

respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions. The formulations to to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e g* , films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. For example, NOVX activity includes growth and differentiation, antibody production, and tumor growth.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. (See Lam 12 Anticancer Drug Design 145 (1997)).

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, et al., 90 Proc. Natl. Acad. Sci. U.S.A. 6909 (1993); Erb, et al., 91 Proc. Natl. Acad. Sci. U.S.A. 11422 (1994); Zuckermann, et al., 37 J. Med. Chem. 2678 (1994); Cho, et al., 261 Science 1303 (1993); Carrell, et al., 33 Angew. Chem. Int. Ed. Engl. 2059 (1993); Carell, et al., 33 Angew. Chem. Int. Ed. Engl. 2061 (1994); and Gallop, et al., 37 J. Med. Chem. 37: 1233 (1994).

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 13 Biotechniques 412-421(1992)), or on beads (Lam, 354 Nature 82-84 (1991)), on chips (Fodor, 364 Nature 555-556 (1993)), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, et al., 89 Proc. Natl. Acad. Sci. USA 1865-1869 (1992)) or on phage (Scott and Smith, 249 Science 386-390 (1990); Devlin, 249 Science 404-406 (1990); Cwirla, et

al., 87 Proc. Natl. Acad. Sci. U.S.A. 6378-6382 (1990); Felici, 222 J. Mol. Biol. 301-310 (1991); Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the

invention In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i e* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test

compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described above.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e g.*, at physiological conditions for

salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes. include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, et al., 72 Cell 223-232 (1993); Madura, et al., 268 J. Biol. Chem. 12046-12054 (1993); Bartel, et al., 14 Biotechniques 920-924 (1993); Iwabuchi, et al., 8 Oncogene 1693-1696 (1993); and Brent, WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) identify an individual from a minute biological sample (tissue typing); and (*ii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Tissue Typing

The NOVX sequences of the invention can be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 1, 3, 5, 7, 9, 11 or 13 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of

the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. Disorders associated with aberrant NOVX expression of activity include, for example, disorders of olfactory loss, *e g.* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease.

The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.
These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 10, 12, or 14, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent

conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

One agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (*e.g.*, for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds.

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of

indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In one embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Such disorders include for example, disorders of olfactory loss, *e.g* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease.

Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can

be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a NOVX gene; (*ii*) an addition of one or more nucleotides to a NOVX gene; (*iii*) a substitution of one or more nucleotides of a NOVX gene, (*iv*) a chromosomal rearrangement of a NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of a NOVX gene, (*vi*) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (*viii*) a non-wild-type level of a NOVX protein, (*ix*) allelic loss of a NOVX gene, and (*x*) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, et al., 241 Science 1077-1080 (1988); and Nakazawa, et al., 91 Proc. Natl. Acad. Sci USA 360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, et al., 23 Nucl. Acids Res. 23: 675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, et al., 87 Proc. Natl. Acad. Sci. USA 1874-1878 (1990)), transcriptional amplification system (*see,* Kwoh, et al., 86 Proc. Natl. Acad. Sci. USA 1173-1177 (1989)); Qβ Replicase (*see,* Lizardi, et al, 6 BioTechnology 1197 (1998)), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g* , DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. See, e.g., Cronin, et al., 7 Human Mutation 244-255 (1996); Kozal, et al., 2 Nat. Med. 753-759 (1996). For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 74 Proc. Natl. Acad. Sci. USA 560 (1997) or Sanger, 74 Proc. Natl. Acad. Sci. USA 5463 (1997). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, et al., 19 Biotechniques 448 (1995)), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, et al., 36 Adv. Chromatography 127-162 (1996); and Griffin, et al., 38 Appl. Biochem. Biotechnol. 47-159 (1993)).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, et al., 230 Science 1242 (1985). In general, the art technique

of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, et al., 85 Proc. Natl. Acad. Sci. USA 4397 (1988); Saleeba, et al., 217 Methods Enzymol. 286-295 (1992). In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, et al., 15 Carcinogenesis 1657-1662 (1994). According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, et al., 86 Proc. Natl. Acad. Sci. USA 2766 (1989); Cotton, 285 Mutat. Res. 125-144 (1993); Hayashi, 9 Genet. Anal. Tech. Appl. 73-79 (1992). Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex

analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen, et al., 7 Trends Genet. 7: 5 (1991).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers, et al., 313 Nature 495 (1985). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 265 Biophys. Chem. 12753 (1987).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, et al., 324 Nature 163 (1986); Saiki, et al., 86 Proc. Natl. Acad. Sci. USA 6230 (1989). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, et al., 17 Nucl. Acids Res. 2437-2448 (1989)) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (see, e.g., Prossner, 11 Tibtech. 11: 238 (1993)). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini, et al., 6 Mol. Cell Probes 1 (1992). It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 88 Proc. Natl. Acad. Sci. USA 189 (1991). In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein,

which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g* disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease). In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g* , drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 23 Clin. Exp. Pharmacol. Physiol., 983-985 (1996); Linder, 43 Clin. Chem., 43: 254-266 (1997). In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is

hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting

decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g* , an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease

expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. Disorders associated with aberrant NOVX expression include, for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (see, e.g., Capecchi, 244 Science 1288-1292 (1989)); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an

aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.,* by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.,* Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual

afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable in situations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated). Another example of such a situation is where the subject has an immunodeficiency disease (*e.g.*, AIDS).

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is

depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1.: Method of Identifying the Nucleic Acids Encoding the G-Protein Coupled Receptors.

Novel nucleic acid sequences were identified by TblastN using CuraGen Corporation's sequence file run against the Genomic Daily Files made available by GenBank. The nucleic acids were further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein.

Example 2: Quantitative Expression Analysis of NOV1

RTQ-PCR Panel Ag431 Description:

As shown in Table 26 below, this 96 well plate (2 control wells, 94 test samples) panel and its variants (Panel 1) are composed of RNA/cDNA isolated from various human cell lines

that have been established from human malignant tissues (Tumors). These cell lines have been extensively characterized by investigators in both academia and the commercial sector regarding their tumorgenicity, metastatic potential, drug resistance, invasive potential and other cancer-related properties. They serve as suitable tools for pre-clinical evaluation of anti-cancer agents and promising therapeutic strategies. RNA from these various human cancer cell lines was isolated by and procured from the Developmental Therapeutic Branch (DTB) of the National Cancer Institute (USA). Basic information regarding their biological behavior, gene expression, and resistance to various cytotoxic agents are known in the art. In addition, RNA/cDNA was obtained from various human tissues derived from human autopsies performed on deceased elderly people or sudden death victims (accidents, etc.). These tissue were ascertained to be free of disease and were purchased from various high quality commercial sources such as Clontech, Inc., Research Genetics , and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electrophoresis using 28s and 18s ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the presence of low molecular weight RNAs indicative of degradation products. Samples are quality controlled for genomic DNA contamination by reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Methods:

The quantitative expression of various clones was assessed in about 41 normal and about 55 tumor samples by real time quantitative PCR (TaqMan®) performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. See Table 24.

First, 96 RNA samples were normalized to β-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48^{0}C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50^{0}C; 10 min. at 95^{0}C; 15 sec. at 95^{0}C/1 min. at 60^{0}C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration

between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their β-actin / GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5` G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below, Table 23) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

TaqMan oligo set Ag431 for the NOV1 gene (*i.e.*, AL135841_B) include the forward, probe, and reverse oligomers shown below:

Table 25.
Gene: AL135841_B
Probe Name: Ag431

Primers	Sequences	Length	Start Position
Forward	5'-AGTCACTTCACCTGCAAGATCCT-3' (SEQ ID NO:25)	23	581
Probe	TET-5'-CCGCATGCCAGCTTCAGCACTG-3'-TAMRA (SEQ ID NO:26)	22	
Reverse	5'-CTTCGCTGACCGACGTGTT-3' (SEQ ID NO:27)	19	629

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (SEQX-

specific and another gene-specific probe multiplexed with the SEQX probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/μl RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute. The results are shown below in Table 26:

TABLE 26

Tissue_Name/Run_Name	1.3Dtm3630t_ag431	Tissue_Name/Run_Name	2Dtm3631t_ag431
Liver adenocarcinoma	3.74	Normal Colon GENPAK 061003	6.93
Heart (fetal)	0	83219 CC Well to Mod Diff (ODO3866)	4.84
Pancreas	0	83220 CC NAT (ODO3866)	0
Pancreatic ca. CAPAN 2	6.75	83221 CC Gr.2 rectosigmoid (ODO3868)	5.4
Adrenal gland	0	83222 CC NAT (ODO3868)	6.61
Thyroid	0	83235 CC Mod Diff (ODO3920)	2.88
Salivary gland	0	83236 CC NAT (ODO3920)	3.79
Pituitary gland	0	83237 CC Gr.2 ascend colon (ODO3921)	6.38
Brain (fetal)	14.66	83238 CC NAT (ODO3921)	6.98
Brain (whole)	22.38	83241 CC from Partial Hepatectomy (ODO4309)	0
Brain (amygdala)	39.5	83242 Liver NAT (ODO4309)	4.18
Brain (cerebellum)	26.61	87472 Colon mets to lung (OD04451-01)	30.99
Brain (hippocampus)	100	87473 Lung NAT (OD04451-02)	2.24
Brain (thalamus)	8.54	Normal Prostate Clontech A+ 6546-1	12.16
Cerebral Cortex	57.83	84140 Prostate Cancer (OD04410)	10.37
Spinal cord	8.96	84141 Prostate NAT (OD04410)	23.65
CNS ca. (glio/astro) U87-MG	0	87073 Prostate Cancer (OD04720-01)	23.33
CNS ca. (glio/astro) U-118-MG	0	87074 Prostate NAT (OD04720-02)	26.61
CNS ca. (astro) SW1783	0	Normal Lung GENPAK 061010	2.05
CNS ca.* (neuro; met) SK-N-AS	11.99	83239 Lung Met to Muscle (ODO4286)	0
CNS ca. (astro) SF-539	3.82	83240 Muscle NAT (ODO4286)	2.03
CNS ca. (astro) SNB-75	6.93	84136 Lung Malignant Cancer (OD03126)	6.75
CNS ca. (glio) SNB-19	0	84137 Lung NAT (OD03126)	14.36
CNS ca. (glio) U251	7.69	84871 Lung Cancer (OD04404)	0
CNS ca. (glio) SF-295	13.97	84872 Lung NAT (OD04404)	3.15
Heart	4.58	84875 Lung Cancer (OD04565)	0
Skeletal muscle	10.08	85950 Lung Cancer (OD04237-01)	1.76
Bone marrow	0	85970 Lung NAT (OD04237-02)	0
Thymus	0	83255 Ocular Mel Met to Liver (ODO4310)	0
Spleen	5.33	83256 Liver NAT (ODO4310)	0

Lymph node	4.74
Colorectal	30.57
Stomach	0
Small intestine	0
Colon ca. SW480	8.3
Colon ca.* (SW480 met)SW620	0
Colon ca. HT29	0
Colon ca. HCT-116	0
Colon ca. CaCo-2	0
83219 CC Well to Mod Diff (ODO3866)	0
Colon ca. HCC-2998	7.75
Gastric ca.* (liver met) NCI-N87	15.39
Bladder	0
Trachea	0
Kidney	7.54
Kidney (fetal)	4.7
Renal ca. 786-0	0
Renal ca. A498	0
Renal ca. RXF 393	0
Renal ca. ACHN	0
Renal ca. UO-31	0
Renal ca. TK-10	0
Liver	0
Liver (fetal)	0
Liver ca. (hepatoblast) HepG2	0
Lung	0
Lung (fetal)	18.56
Lung ca. (small cell) LX-1	15.28
Lung ca. (small cell) NCI-H69	0
Lung ca. (s.cell var.) SHP-77	5.11
Lung ca. (large cell)NCI-H460	0
Lung ca. (non-sm. cell) A549	0
Lung ca. (non-s.cell) NCI-H23	4.21
Lung ca (non-s.cell) HOP-62	0
Lung ca. (non-s.cl) NCI-H522	4.67
Lung ca. (squam.) SW 900	5.08
Lung ca. (squam.) NCI-H596	0
Mammary gland	7.08
Breast ca.* (pl. effusion) MCF-7	0
Breast ca.* (pl.ef) MDA-MB-231	4.21
Breast ca.* (pl. effusion) T47D	0
Breast ca. BT-549	0

84139 Melanoma Mets to Lung (OD04321)	0
84138 Lung NAT (OD04321)	8.42
Normal Kidney GENPAK 061008	8.42
83786 Kidney Ca, Nuclear grade 2 (OD04338)	3.06
83787 Kidney NAT (OD04338)	0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0
83789 Kidney NAT (OD04339)	10.37
83790 Kidney Ca, Clear cell type (OD04340)	0
83791 Kidney NAT (OD04340)	0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0
83793 Kidney NAT (OD04348)	9.02
87474 Kidney Cancer (OD04622-01)	3.59
87475 Kidney NAT (OD04622-03)	0
85973 Kidney Cancer (OD04450-01)	0
85974 Kidney NAT (OD04450-03)	17.08
Kidney Cancer Clontech 8120607	0
Kidney NAT Clontech 8120608	0
Kidney Cancer Clontech 8120613	0
Kidney NAT Clontech 8120614	0
Kidney Cancer Clontech 9010320	0
Kidney NAT Clontech 9010321	70.22
Normal Uterus GENPAK 061018	6.08
Uterus Cancer GENPAK 064011	16.49
Normal Thyroid Clontech A+ 6570-1	0
Thyroid Cancer GENPAK 064010	0
Thyroid Cancer INVITROGEN A302152	0
Thyroid NAT INVITROGEN A302153	6.34
Normal Breast GENPAK 061019	3.12
84877 Breast Cancer (OD04566)	100
85975 Breast Cancer (OD04590-01)	9.41
85976 Breast Cancer Mets (OD04590-03)	0
87070 Breast Cancer Metastasis (OD04655-05)	49.31
GENPAK Breast Cancer 064006	31.86
Breast Cancer Clontech 9100266	13.4
Breast NAT Clontech 9100265	4.36
Breast Cancer INVITROGEN A209073	4.74
Breast NAT INVITROGEN A2090734	9.21
Normal Liver GENPAK 061009	0
Liver Cancer GENPAK 064003	2.5
Liver Cancer Research Genetics RNA 1025	0
Liver Cancer Research Genetics RNA 1026	0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	7.48

Breast ca.	MDA-N	11.34	Paired Liver Tissue Research Genetics RNA 6004-N	8.13
Ovary		0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0
Ovarian ca.	OVCAR-3	0	Paired Liver Tissue Research Genetics RNA 6005-N	0
Ovarian ca.	OVCAR-4	0	Normal Bladder GENPAK 061001	12.07
Ovarian ca.	OVCAR-5	0	Bladder Cancer Research Genetics RNA 1023	9.34
Ovarian ca.	OVCAR-8	0	Bladder Cancer INVITROGEN A302173	9.47
Ovarian ca.	IGROV-1	0	87071 Bladder Cancer (OD04718-01)	2.68
Ovarian ca.* (ascites)	SK-OV-3	0	87072 Bladder Normal Adjacent (OD04718-03)	10.73
Uterus		0	Normal Ovary Res. Gen.	0
Plancenta		0	Ovarian Cancer GENPAK 064008	6.65
Prostate		0	87492 Ovary Cancer (OD04768-07)	0
Prostate ca.* (bone met)	PC-3	0	87493 Ovary NAT (OD04768-08)	0
Testis		13.87	Normal Stomach GENPAK 061017	16.38
Melanoma	Hs688(A).T	0	NAT Stomach Clontech 9060359	0
Melanoma* (met)	Hs688(B).T	0	Gastric Cancer Clontech 9060395	9.67
Melanoma	UACC-62	4.27	NAT Stomach Clontech 9060394	1.95
Melanoma	M14	0	Gastric Cancer Clontech 9060397	1.92
Melanoma	LOX IMVI	0	NAT Stomach Clontech 9060396	0
Melanoma* (met)	SK-MEL-5	0	Gastric Cancer GENPAK 064005	1.9
Adipose		0		

In Table 26, the following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var= small cell variant,
non-s = non-sm =non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

These results are summarized below:

TABLE 27

NOVX	Internal Accession Number	Results
NOV1	AL135841_B	Ag431, potential utilities for breast cancer, several cancer in panel 2 and couple of cell lines in panel 1

Example 3.

The DNA and protein sequences for the novel single nucleotide polymorphic variants of the Olfactory Receptor-like NOV4 gene of CuraGen Acc. No. CG54212-01 are reported in Tables 15 and 16. Variants are reported individually but any combination of all or a select subset of variants are also included. In Tables 15 and 16, the positions of the variant bases and the variant amino acid residues are underlined. In summary, there is one variant reported in Tables 15 and 16. Variant 13019736 is a T to C SNP at 236 bp of the nucleotide sequence that results in a Tyr to His change at amino acid 60 of protein sequence.

The association of the novel single nucleotide polymorphic variant in Table 15 with specific phenotypic traits is reported in Table 17.

The serum levels of gamma-glutamyl transpeptidase are significantly associated with this variant, with a statistical significance level of 0.0001. The presence of this variant allele is associated with a decrease in serum gamma-glutamyl transpeptidase levels of 0.4 standard deviations below the mean level in the sampled population. Elevated serum levels of gamma-glutamyl transpeptidase are risk factors for hepatic damage and liver disease, therefore the SNP reported here may be a specific marker for a statistically significant decreased risk of liver disease. The Olfactory-receptor-like protein of the invention is a novel target for pharmaceutical and other therapeutic interventions important in liver disease, and has additional utility as a diagnostic marker.

The serum levels of calcium and measures of regional bone density are also significantly associated with the novel NOV4 variant depicted in Table 16, with a statistical significance level of 0.0005 and 0.002, respectively. The presence of this variant allele is associated with an increase in bone density of 0.4 standard deviations above the mean level in the sampled population, and a decrease in serum calcium of 0.4 standard deviations below the mean level in the sampled population. Bone density and serum calcium levels are risk factors for osteoporosis as well as other bone and skeletal disorders, therefore the SNP reported here may be a specific marker for a statistically significant altered risk of osteoporosis and other bone diseases. The Olfactory-receptor-like protein of the invention is a novel target for pharmaceutical and other therapeutic interventions important in bone disease, and has additional utility as a diagnostic marker.

Example 4.

The DNA and protein sequences for the novel single nucleotide polymorphic variants of the Olfactory-like NOV6 gene of CuraGen Acc. No. CG53482-01 are reported in Tables 20 and 21. Variants are reported individually but any combination of all or a select subset of variants are also included. In Tables 20 and 21, the positions of the variant bases and the variant amino acid residues are underlined. In summary, there is one variant reported in Tables 20 and 21. Variant 13373788 is a T to C SNP at 278 bp of the nucleotide sequence that results in no change in the protein coding sequence (silent).

The association of the novel single nucleotide polymorphic variant in Tables 20 and 21 with specific phenotypic traits is reported in Table 22. The serum levels of apolipoprotein(a) are significantly associated with this variant, with a statistical significance level of 0.0001. The presence of this variant allele is associated with an increase in serum apolipoprotein(a) levels of 0.4 standard deviations above the mean level in the sampled population.

Elevated serum levels of apolipoprotein(a) are risk factors for coronary heart disease and carotid atherosclerosis, therefore the SNP reported here may be a specific marker for a statistically significant increased risk of cardiovascular disease. The Olfactory-receptor-like protein of the invention is a novel target for pharmaceutical and other therapeutic interventions important in cardiovascular disease, and has additional utility as a diagnostic marker.

Example 5.

ClustalW analyses of the NOVX sequences were performed as shown below.

Table 28. ClustalW comparison of NOV1, NOV2, NOV4 and NOV5.

```
NOV1   MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSV  50
NOV2   MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSV  50
NOV4   MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSV  50
NOV5   MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSV  50

NOV1   LDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQ  100
NOV2   LDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQ  100
NOV4   LDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQ  100
NOV5   LDIHLHTPVHFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQ  100

NOV1   MCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWV  150
NOV2   MCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWV  150
NOV4   MCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWV  150
NOV5   MCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWV  150
```

```
NOV1   LCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLL 200
NOV2   LCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLL 200
NOV4   LCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLL 200
NOV5   LCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLL 200

NOV1   AGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLL 250
NOV2   AGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLL 250
NOV4   AGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLL 250
NOV5   AGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLL 250

NOV1   FYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300
NOV2   FYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300
NOV4   FYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300
NOV5   FYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA 300

NOV1   ARKVWGRSRASR 312
NOV2   ARKVWGRSRASR 312
NOV4   ARKVWGRSRASR 312
NOV5   ARKVWGRSRASR 312
```

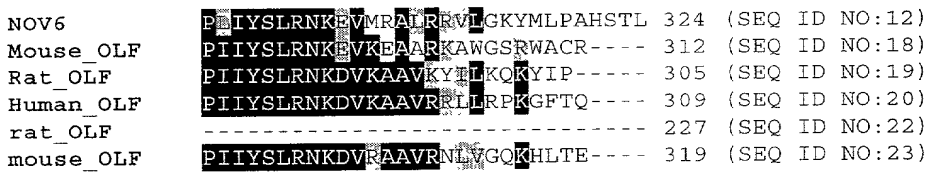
Table 29. ClustalW analysis of NOV4 with GPCR family members.

```
NOV4          -MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVS  49
Mouse_OLF     -MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGNTAIVAVS  49
Rat_OLF       -------------LLLGLSGYPKTEILYFVIVLVMYLVIHTGNGVLIIAS  37
Human_OLF     ----------MGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVT  40
rat_OLF       ----------------------------------------------------  1
mouse_OLF     MDRSNETAPLSGFILLGLSAHPKLEKTFFVLILMMYLVILLGNGVLILVS  50

NOV4          VLDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAL  99
Mouse_OLF     MLDARLHTPMYFFLGNLSILDICYTSTFVPLMLVHLLSSRKTISFIGCAV  99
Rat_OLF       IFDSHLHTPMYFFLGNLSFLDICYTTSSVPSTLVSLISKKRNISFSGCTV  87
Human_OLF     ILDSRLHTPMYFFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAV  90
rat_OLF       --------------SNLSFLDICYTTSSVPLILGSFLTPRKTISFSGCAV  36
mouse_OLF     ILDSHLHTPMYFFLGNLSFLDICYTTSSVPLILDSFLTPRKTISFSGCAV 100

NOV4          QMCLSLSTGSTECLLLALTAYDRYLAICQPLRYHVLMSHRICVLLMGAAW 149
Mouse_OLF     QMCLSLSTGSTECLLLAVMAYDRYLAICQPLRYPVLMSHRICLMLAGASW 149
Rat_OLF       QMFVGFAMGSTECLLLGMMAFDRYVAICNPLRYSVIMSKEVYVSMASASW 137
Human_OLF     QMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVIMSKAAYMPMAASSW 140
rat_OLF       QMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMSKAVYVPMATGSW  86
mouse_OLF     QMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMNKAAYVPMAASSW 150

NOV4          VLCLLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFL 199
Mouse_OLF     VLCLFKSVAETVIAMRLPFCGHHVIRHFTCEILAVLKLTCGDTSVSDAFL 199
Rat_OLF       FSGGINSVVQTSLAMRLPFCGNNVINHFTCEVLAVLKLACADISLNIVTM 187
Human_OLF     AIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISM 190
rat_OLF       AAGIAASLVQTSLAMRLPFCGDNVINHFTCEILAVLKLACADISINIISM 136
mouse_OLF     AGGITNSVVQTSLAMRLPFCGDNVINHFTCEILAVLKLACADISINVISM 200

NOV4          LAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVL 249
Mouse_OLF     LVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFSTCSAHLAVVL 249
Rat_OLF       VISNMAFLVLPLLLIFFSYVLILYTILRMNSASGRRKAFSTCSAHLTVVV 237
Human_OLF     EVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVI 240
rat_OLF       GVTNVIFLGVPVLFISFSYIFILSTILRIPSAEGRKKAFSTCSAHLTVVI 186
mouse_OLF     VVANMIFLAVPVLFIFVSYVFILVTILRIPSAEGRKKAFSTCSAHLTVVL 250
```

```
NOV4          IFYGTIIFMYLKPKSKE---A---HISDEVFTVLVAMVTTMLNPTIYSLR 293
Mouse_OLF     IFYSTIIFMYMKPKSKE---A---RISDQVFTVLVAVVTPMLNPIIYSLR 293
Rat_OLF       IFYGTIFSMYAKPKSQDLTGKDKFQTSDKIISLFYGVVTPMLNPIIYSLR 287
Human_OLF     VFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLR 290
rat_OLF       VFYGTILFMYGKPKSKDPLGADKQDPADKLISLFYGVLTPM--------- 227
mouse_OLF     VFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGVVTPMLNPIIYSLR 300

NOV4          NKEVKEAARKVWGRSRASR 312 (SEQ ID NO:8)
Mouse_OLF     NKEVKEAARKAWGSRWACR 312 (SEQ ID NO:18)
Rat_OLF       NKDVKAAVKYTLKQKYIP- 305 (SEQ ID NO:19)
Human_OLF     NKDVKAAVRRLLRPKGFTQ 309 (SEQ ID NO:20)
rat_OLF       ------------------- 227 (SEQ ID NO:22)
mouse_OLF     NKDVRAAVRNLVGQKHLTE 319 (SEQ ID NO:23)
```

Table 30. ClustalW analysis of NOV6 with GPCR family members.

```
NOV6          MELWNYHSMELWNFTLGSGFILVGILNDSGSPLLLCATITLLYLLALLSN 50
Mouse_OLF     --------MEPSNRTAVSEFVLKGFSGYPALERLLFPLCSVMYLVTLLGN 42
Rat_OLF       ----------------------LLLGLSGYPKTEILYFVIVLVMYLVIHTGN 30
Human_OLF     -----------------MGFVLLRLSAHPELEKTFFVLILLMYLVILLGN 33
rat_OLF       ---------------------------------------------------- 1
mouse_OLF     -------MDRSNETAPLSGFTLLGLSAHPKLEKTFFVLILVMYLVILLGN 43

NOV6          GLLLLAITMEARLHMPMYLLLGQLSLMDLLFTSVVTPKALADFLRRENTI 100
Mouse_OLF     TAIVAVSMLDARLHTPMYFFLGNLSILDICYTSTFVPLMLVHLLSSRKTI 92
Rat_OLF       GVLITASIFDSHLHTPMYFFLGNLSFLDICYTTSSVPSTLVSLISKKRNI 80
Human_OLF     GVLILVTILDSRLHTPMYFFLGNLSFLDICFTTSSVPLVLDSFLTPQETI 83
rat_OLF       ----------------------SNLSFLDICYTTSSVPLILGSFLTPRKTI 29
mouse_OLF     GVLILVSILDSHLHTPMYFFLGNLSFLDICYTTSSVPLILDSFLTPRKTI 93

NOV6          SFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACW 150
Mouse_OLF     SFTGCAVQMCLSLSTGSTECLLLAVMAYDRYLAICQPLRYPVLMSHRLCL 142
Rat_OLF       SFSGCTVQMFVGFAMGSTECLLLGMMAFDRYVAICNPLRYSVIMSKEVYV 130
Human_OLF     SFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVIMSKAAYM 133
rat_OLF       SFSGCAVQMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMSKAVYV 79
mouse_OLF     SFSGCAVQMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMNKAAYV 143

NOV6          LMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLACADT 200
Mouse_OLF     MLAGASWVLCLFKSVAETVLAMRLPFCGHHVIRHFTCEILAVLKLTCGDT 192
Rat_OLF       SMASASWFSCGINSVVQTSLAMRLPFCGNNVINHFTCEVLAVLKLACADI 180
Human_OLF     PMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADI 183
rat_OLF       PMATGSWAAGIAASLVQTSLAMRLPFCGDNVINHFTCEILAVLKLACADI 129
mouse_OLF     PMAASSWAGGITNSVVQTSLAMRLPFCGDNVINHFTCEILAVLKLACADI 193

NOV6          SRYELMVYVMGVTFLIPSLAATLASYTQILLTVLHMPSNEGRKKALVTCS 250
Mouse_OLF     SVSDAFTLVGAILLLPIPLTLICLSYMLILATILRVPSATGRSKAFSTCS 242
Rat_OLF       SLNIVTMVISNMAFLVLPLLLIFFSYVLILYTILRMNSASGRRKAFSTCS 230
Human_OLF     SINVISMEVTNVIFLGVPVLFISFSYVFIFTTILRIPSAEGRKKVFSTCS 233
rat_OLF       SINIISMGVTNVIFLGVPVLFISFSYIFILSTILRIPSAEGRKKAFSTCS 179
mouse_OLF     SINVISMVVANMIFLAVPVLFIFVSYVFILVTILRIPSAEGRKKAFSTCS 243

NOV6          SHLTVVGMFYGAATFMYVLPSSFHS------TRQDNIISVFYTIVTPALN 294
Mouse_OLF     AHLAVVLLFYSTIIFMYMKPKSKE---A---RISDQVFTVLVAVVTPMLN 286
Rat_OLF       AHLTVVVIFYGTIFSMYAKPKSQDLTGKDKFQTSDKIISLFYGVVTPMLN 280
Human_OLF     AHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLN 283
rat_OLF       AHLTVVIVFYGTILFMYGKPKSKDPLGADKQDPADKLISLFYGVLTPM-- 227
mouse_OLF     AHLTVVIVFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGVVTPMLN 293
```



OTHER EMBODIMENTS

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.